AD-A267 495

VIENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1993	~~MASTER'S~~/DISSERTATION

4. TITLE AND SUBTITLE
A Polnomial Primal-Dual Interior Point Method for Convex Programming with Quadratic Constraints

5. FUNDING NUMBERS

6. AUTHOR(S)
Lee James Lehmkuhl

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Student Attending: The George Washington Univ

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/CI/CIA- 93-015D

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTMENT OF THE AIR FORCE
AFIT/CI
2950 P STREET
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
AUG 6 1993
S C D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release IAW 190-1
Distribution Unlimited
MICHAEL M. BRICKER, SMSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

93 8 05 116

93-18068

14. SUBJECT TERMS

15. NUMBER OF PAGES
124

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

ABSTRACT OF DISSERTATION

A Polynomial Primal-Dual Interior Point Method
for Convex Programming With Quadratic Constraints

by
Lee James Lehmkuhl
Captain, United States Air Force
Doctor of Science in Operations Research
The George Washington University

This study involves the solution of a convex nonlinear programming problem using a primal-dual interior point algorithm. The problem is to minimize $b^T x$ subject to $g_i(x) \geq 0$, i=1 to m, where each $g_i(x)$ is a concave quadratic function. We specify certain common regularity conditions to guarantee the existence of a solution x^* and a vector of KKT multipliers $u^* \geq 0$ such that (x^*, u^*) solves the associated Wolfe dual. The algorithm is motivated by the following optimality conditions, given a fixed parameter $\mu > 0$, for x and u such that $g_i(x) > 0$ and $u_i > 0$, i = 1, . . ., m:

$$\nabla L(x,u) = b - \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$$

$$u_i g_i(x) - \mu = 0, \; i = 1, \ldots, m.$$

The algorithm uses Newton's method to approximately solve these equations for a decreasing sequence $\{\mu^k\}$ where $\mu^k \downarrow 0$ as $k \to +\infty$. A step size procedure maintains feasibility and seeks to decrease some merit function.

The Newton direction obtained has several interesting features. The direction in the primal variable x is closely related to the Newton direction generated by a Sequential Unconstrained Minimization Technique employing the logarithmic barrier function (SUMT). In fact, the primal direction can be viewed as a SUMT Newton direction with a perturbed barrier function Hessian matrix. The perturbation depends on the degree to which $u_i g_i(x) - \mu = 0$, or perturbed complementary slackness (PCS), is violated for each i. If the deviation from PCS is assumed to be small, the primal direction is close enough to the SUMT Newton direction that it may be used as the search direction in a SUMT algorithm. The resulting primal variable algorithm retains the polynomial computational complexity shown for SUMT.

The Newton direction in the dual variables generated by the primal-dual algorithm is also studied. A modification to the dual step size and direction allows the progress in reducing PCS at each iteration to be calculated and therefore controlled. This control allows the deviation from PCS to be kept small

throughout the primal-dual algorithm's progress, and thus the polynomial complexity shown for the primal variable algorithm discussed above follows for the primal-dual algorithm.

(Completed in 1993; 132 pages.)

BIBLIOGRAPHY

[1] Anstreicher, K.. 1990. "On long step path following and SUMT for linear and quadratic programming," Technical Paper, Department of Operations Research, Yale University, New Haven, Connecticut 06520.

[2] Anstreicher, K., D. den Hertog, C. Roos, and T. Terlaky. 1990. "A long step barrier method for convex quadratic programming," Report 90-53, Faculty of Mathematics and Informatics/Computer Science, Delft University of Technology, Delft, Netherlands.

[3] Bazaraa, M. S., J. J. Jarvis, and H. D. Sherali. 1990. Linear Programming and Network Flows, John Wiley and Sons, New York.

[4] Charnes, A., and K. O. Kortanek. 1965. "Extreme Point Solutions in Mathematical Programming: An Opposite Sign Algorithm," Systems Research Memorandum No. 129, Northwestern University, Evanston, Illinois.

[5] den Hertog, D., C. Roos, and T. Terlaky. 1990. "On the classical logarithmic barrier function method for a class of smooth convex programming problems," Report 90-28, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands.

[6] Fiacco, A. V.. 1979. "Barrier Methods for Nonlinear Programming." In Operations Research Support Methodology, (Industrial Engineering Series Volume 2), A. G. Holzman, (Ed.), Marcel Dekker, Inc., New York, pp. 377-440.

[7] Fiacco, A. V., and G. P. McCormick. 1968, 1990. Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, New York. An unabridged, corrected re-publication was published by SIAM in the series Classics in Applied Mathematics (1990).

[8] Gonzaga, C. C.. 1992. "Path following methods for linear programming," SIAM Review, Vol. 34, No. 2, pp. 167-224.

[9] Huang, S., and K. O. Kortanek. 1991. "On a potential reduction algorithm for LP with simultaneous primal-dual updating," Operations Research Letters 10, pp. 501-507.

[10] Huard, P.. 1967. "Resolution of mathematical programming with nonlinear constraints by the method of centers," in Nonlinear Programming, J. Abadie (Ed.), North Holland, Amsterdam.

[11] Jarre, F.. 1991. "On the convergence of the method of analytical centers when applied to convex quadratic programs," Mathematical Programming 49, pp. 341-358.

[12] Karmakar, N.. 1984. "A New Polynomial-time Algorithm for Linear Programming," Combinatoria, 4, pp. 373-395.

[13] Khachian, L. G.. 1979. "A Polynomial Algorithm In Linear Programming," Soviet Mathematics Doklady, 20, pp. 191-194.

[14] Klee, V., and G. J. Minty. 1972. "How Good is the Simplex Algorithm?" in Inequalities III, O. Shisha, (Ed.), Academic, New York, pp. 159-175.

[15] McCormick, G. P.. 1983. Nonlinear Programming: Theory, Algorithms, and Applications, John Wiley and Sons, New York.

[16] McCormick, G. P.. 1991a. "Resolving the shell dual with a nonlinear primal-dual algorithm," T-545/91, Department of Operations Research, The George Washington University, Washington, DC 20052.

[17] McCormick, G. P.. 1991b. "The superlinear convergence of a nonlinear primal-dual algorithm," T-550/91, Department of Operations Research, The George Washington University, Washington, DC 20052.

[18] Mehrotra, S.. 1990. "On finding a vertex solution using interior point methods," Technical Report 89-22, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208.

[19] Mehrotra, S., and J. Sun. 1990. "An algorithm for convex quadratic programming that requires $O(n^{3.5}L)$ arithmetic operations. Mathematics of Operations Research, Vol. 15, No. 2, pp. 342-363.

[20] Papadimitriou, C. H., and K. Steiglitz. 1982. Combinatorial Optimization: Algorithms and Complexity, Prentice-Hall, Inc., Engelwood Cliffs, New Jersey.

[21] Powell, M. J. D.. 1991. "On the number of iterations of Karmarkar's algorithm for linear programming," Numerical Analysis Report 23, Department of Applied Mathmatics and Theoretical Physics, University of Cambridge, Cambridge, England.

[22] Wright, M. H.. 1991. "Interior methods for constrained optimization," in Acta Numerica, Vol. 1, A. Iserles (Ed.), Cambridge University Press, New York, pp. 341-407.

ABSTRACT OF DISSERTATION

A Polynomial Primal-Dual Interior Point Method
for Convex Programming With Quadratic Constraints

by
Lee James Lehmkuhl
Captain, United States Air Force
Doctor of Science in Operations Research
The George Washington University

This study involves the solution of a convex nonlinear programming problem using a primal-dual interior point algorithm. The problem is to minimize $b^T x$ subject to $g_i(x) \geq 0$, i=1 to m, where each $g_i(x)$ is a concave quadratic function. We specify certain common regularity conditions to guarantee the existence of a solution x^* and a vector of KKT multipliers $u^* \geq 0$ such that (x^*, u^*) solves the associated Wolfe dual. The algorithm is motivated by the following optimality conditions, given a fixed parameter $\mu > 0$, for x and u such that $g_i(x) > 0$ and $u_i > 0$, i = 1, . . ., m:

$$\nabla L(x,u) = b - \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$$

$$u_i g_i(x) - \mu = 0, \; i = 1, \ldots, m.$$

The algorithm uses Newton's method to approximately solve these equations for a decreasing sequence $\{\mu^k\}$ where $\mu^k \downarrow 0$ as $k \rightarrow +\infty$. A step size procedure maintains feasibility and seeks to decrease some merit function.

The Newton direction obtained has several interesting features. The direction in the primal variable x is closely related to the Newton direction generated by a Sequential Unconstrained Minimization Technique employing the logarithmic barrier function (SUMT). In fact, the primal direction can be viewed as a SUMT Newton direction with a perturbed barrier function Hessian matrix. The perturbation depends on the degree to which $u_i g_i(x) - \mu = 0$, or perturbed complementary slackness (PCS), is violated for each i. If the deviation from PCS is assumed to be small, the primal direction is close enough to the SUMT Newton direction that it may be used as the search direction in a SUMT algorithm. The resulting primal variable algorithm retains the polynomial computational complexity shown for SUMT.

The Newton direction in the dual variables generated by the primal-dual algorithm is also studied. A modification to the dual step size and direction allows the progress in reducing PCS at each iteration to be calculated and therefore controlled. This control allows the deviation from PCS to be kept small

throughout the primal-dual algorithm's progress, and thus the polynomial complexity shown for the primal variable algorithm discussed above follows for the primal-dual algorithm.

(Completed in 1993; 132 pages.)

BIBLIOGRAPHY

[1] Anstreicher, K.. 1990. "On long step path following and SUMT for linear and quadratic programming," Technical Paper, Department of Operations Research, Yale University, New Haven, Connecticut 06520.

[2] Anstreicher, K., D. den Hertog, C. Roos, and T. Terlaky. 1990. "A long step barrier method for convex quadratic programming," Report 90-53, Faculty of Mathematics and Informatics/Computer Science, Delft University of Technology, Delft, Netherlands.

[3] Bazaraa, M. S., J. J. Jarvis, and H. D. Sherali. 1990. Linear Programming and Network Flows, John Wiley and Sons, New York.

[4] Charnes, A., and K. O. Kortanek. 1965. "Extreme Point Solutions in Mathematical Programming: An Opposite Sign Algorithm," Systems Research Memorandum No. 129, Northwestern University, Evanston, Illinois.

[5] den Hertog, D., C. Roos, and T. Terlaky. 1990. "On the classical logarithmic barrier function method for a class of smooth convex programming problems," Report 90-28, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands.

[6] Fiacco, A. V.. 1979. "Barrier Methods for Nonlinear Programming." In Operations Research Support Methodology, (Industrial Engineering Series Volume 2), A. G. Holzman, (Ed.), Marcel Dekker, Inc., New York, pp. 377-440.

[7] Fiacco, A. V., and G. P. McCormick. 1968, 1990. Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, New York. An unabridged, corrected re-publication was published by SIAM in the series Classics in Applied Mathematics (1990).

[8] Gonzaga, C. C.. 1992. "Path following methods for linear programming," SIAM Review, Vol. 34, No. 2, pp. 167-224.

[9] Huang, S., and K. O. Kortanek. 1991. "On a potential reduction algorithm for LP with simultaneous primal-dual updating," Operations Research Letters 10, pp. 501-507.

[10] Huard, P.. 1967. "Resolution of mathematical programming with nonlinear constraints by the method of centers," in Nonlinear Programming, J. Abadie (Ed.), North Holland, Amsterdam.

[11] Jarre, F.. 1991. "On the convergence of the method of analytical centers when applied to convex quadratic programs," Mathematical Programming 49, pp. 341-358.

[12] Karmakar, N.. 1984. "A New Polynomial-time Algorithm for Linear Programming," Combinatoria, 4, pp. 373-395.

[13] Khachian, L. G.. 1979. "A Polynomial Algorithm In Linear Programming," Soviet Mathematics Doklady, 20, pp. 191-194.

[14] Klee, V., and G. J. Minty. 1972. "How Good is the Simplex Algorithm?" in Inequalities III, O. Shisha, (Ed.), Academic, New York, pp. 159-175.

[15] McCormick, G. P.. 1983. Nonlinear Programming: Theory, Algorithms, and Applications, John Wiley and Sons, New York.

[16] McCormick, G. P.. 1991a. "Resolving the shell dual with a nonlinear primal-dual algorithm," T-545/91, Department of Operations Research, The George Washington University, Washington, DC 20052.

A POLYNOMIAL PRIMAL-DUAL INTERIOR POINT METHOD FOR CONVEX PROGRAMMING WITH QUADRATIC CONSTRAINTS

By

Lee James Lehmkuhl

B.S. May 1981, University of Iowa

M.S. December 1986, Air Force Institute of Technology

A Dissertation submitted to

The Faculty of

The School of Engineering and Applied Science of
The George Washington University in partial satisfaction
of the requirements for the degree of Doctor of Science

April 20, 1993

Dissertation directed by

Anthony V. Fiacco

Professor of Operations Research

[17] McCormick, G. P.. 1991b. "The superlinear convergence of a nonlinear primal-dual algorithm," T-550/91, Department of Operations Research, The George Washington University, Washington, DC 20052.

[18] Mehrotra, S . 1990. "On finding a vertex solution using interior point methods," Technical Report 89-22, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208.

[19] Mehrotra, S., and J. Sun. 1990. "An algorithm for convex quadratic programming that requires $O(n^{3.5}L)$ arithmetic operations. Mathematics of Operations Research, Vol. 15, No. 2, pp. 342-363.

[20] Papadimitriou, C. H., and K. Steiglitz. 1982. Combinatorial Optimization: Algorithms and Complexity, Prentice-Hall, Inc., Engelwood Cliffs, New Jersey.

[21] Powell, M. J. D.. 1991. "On the number of iterations of Karmarkar's algorithm for linear programming," Numerical Analysis Report 23, Department of Applied Mathmatics and Theoretical Physics, University of Cambridge, Cambridge, England.

[22] Wright, M. H.. 1991. "Interior methods for constrained optimization," in Acta Numerica, Vol. 1, A. Iserles (Ed.), Cambridge University Press, New York, pp. 341-407.

ACKNOWLEDGEMENTS

I wish to thank Professor Garth McCormick for originally introducing me to this area of research, and for his confidence in my ability to handle it. I am indebted to Professor Anthony Fiacco for both his willingness, at rather short notice, to serve as my research advisor on an already developing topic, and his timely, accurate corrections and recommendations. I also benefited from two thorough and thoughtful readers, Professors James Falk and Hernán Abeledo, who improved this dissertation with their insightful comments. Thanks are due as well to my committee members, Professors William Marlow, Richard Soland, and Thomas Mazzuchi.

Finally, I thank the United States Air Force, specifically the Air Force Institute of Technology, for sponsoring me in this degree program; and my family, for its support in this endeavor.

Abstract

A POLYNOMIAL PRIMAL-DUAL INTERIOR POINT
METHOD FOR CONVEX PROGRAMMING
WITH QUADRATIC CONSTRAINTS

by Lee James Lehmkuhl

Anthony V. Fiacco, Director of Research

This dissertation involves the solution of a convex nonlinear programming problem using a primal-dual algorithm developed in McCormick(1991a and 1991b). The problem is to minimize $b^{\mathrm{T}}x$ subject to $g_i(x) \geq 0$, i=1 to m, where each $g_i(x)$ is a concave quadratic function. We specify certain common regularity conditions to guarantee the existence of a solution x^* and a vector of KKT multipliers $u^* \geq 0$ such that (x^*, u^*) solves the associated Wolfe dual: maximize $L(x,u) = b^{\mathrm{T}}x - \sum_{i=1}^{m} u_i g_i(x)$ subject to $\nabla_x L(x,u)=0$ and $u \geq 0$. The algorithm is motivated by the following optimality conditions for the problem stated above, given a fixed parameter $\mu > 0$:

$$\nabla_x L(x,u) = 0$$
$$u_i g_i(x) - \mu = 0, \; i=1,\ldots, m.$$

The algorithm uses Newton's method to approximately solve these equations for a decreasing sequence $\{\mu^k\}$ where $\mu^k \downarrow 0$ as $k \rightarrow +\infty$. A step size procedure maintains feasibility and seeks to decrease some merit function. The Newton direction is the

negative inverse of the Jacobian matrix of the above equations multiplied by the vector of the equations.

The Newton direction obtained has several interesting features. The direction in the primal variable x is closely related to the Newton direction generated by a Sequential Unconstrained Minimization Technique employing the logarithmic barrier function (SUMT). In fact, the primal direction can be viewed as a SUMT Newton direction with a perturbed barrier function Hessian matrix. The perturbation, that is, the degree to which the primal direction differs from the SUMT Newton direction depends on the degree to which $u_i g_i(x) - \mu = 0$, or perturbed complementary slackness (PCS), is violated for each i. If the deviation from PCS is assumed to be small, the primal direction is close enough to the SUMT Newton direction that it may be used as the search direction in a SUMT algorithm. The resulting primal variable algorithm retains the polynomial computational complexity shown for SUMT by den Hertog, Roos, and Terlaky(1990). The Newton direction in the dual variables generated by the primal-dual algorithm is also studied. This direction seeks to change each dual variable to adapt to changes in its associated constraint $g_i(x)$ as x changes along the primal Newton direction, and to reduce the violation of PCS at the current value of x. A modification to the dual step size and direction allows the progress in reducing PCS at each iteration to be calculated and therefore controlled. This control allows the deviation from PCS to be kept small throughout the primal-dual algorithm's progress, and thus the polynomial complexity shown for the primal variable algorithm discussed above follows for the primal-dual algorithm. Finally,

the effect of the modification also allows the use of the results in Anstreicher(1990) for quadratic programming with SUMT. These can be applied to show another proof of polynomiality for linear programming with the modified primal-dual algorithm.

TABLE OF CONTENTS

CHAPTER I: INTRODUCTION

I.1 Background

Constrained optimization is the process of minimizing (or maximizing) the value of a function of one or more variables over a set of allowable values for those variables. The set of allowable values is known as the feasible region. This thesis presents a method, or algorithm, of constrained optimization for a particular class of problems. Subsequent chapters contain a rigorous definition of the problem to be solved, as well as coverage of fundamental theory and related research; but for introductory purposes a general, non-rigorous background description is in order. The class of problems considered here requires the minimization of a convex quadratic function defined over E^n, known as the objective function. The feasible region is defined by one or more concave quadratic functions, known as constraints, in the following manner: the feasible region is the set of those values of $x \in E^n$ which cause each constraint to be non-negative. The algorithm moves toward the solution along a series of points in the interior of the feasible region, that is, the subset of the feasible region where the constraints are strictly positive; hence the

term "interior point method". The problem structure provides a convex feasible region, which together with the convex objective function results in a convex quadratic minimization problem. This class of problems is of significant interest--for instance, a constraint on the maximum Euclidian distance between two points may be expressed with a concave quadratic constraint like that mentioned above. So for example, the minimization of the distance between a number of facilities subject to constraints on the facilities locations, perhaps within a certain service area, may be modeled as such a problem.

There are several important classes of problems which conform to this structure. Perhaps the most common and well-known is the linear program, in which the objective function is linear and the constraints are linear affine. Another is the convex quadratic program, with a convex quadratic objective function and linear affine constraints. Such problems occur in practice and require solution. For example, linear programming may be used to plan the employment of a limited arsenal of nuclear weapons to maximize the damage inflicted on a foe, or to minimize friendly casualties for a given level of destruction. Another application is the allocation of limited resources (labor, parts, equipment, and so on) within a factory to maximize profit.

Much research has gone into developing algorithms to solve problems of this type. Some specialized algorithms apply just to the linear program, and others to the convex quadratic program. Initially, the main goal in algorithm development was to guarantee convergence to an optimal value. Then, the speed of convergence and computational considerations came into focus. Some algorithms which perform

well in practice can be shown in a worst case scenario to require a prohibitive amount of computation effort, while others which possess a more reasonable upper bound on computational effort may converge so slowly in practice that they prove useless for solving actual problems. One can thus see the development of two measures of algorithmic performance. One is the algorithm's speed, in terms of computer time, in solving a wide variety of test problems. Another is the theoretical bound on the number of computations required in the worst case to solve a general problem of a given size. (The measure of problem size will be discussed later.) In this paper, the algorithm presented will be analyzed for convergence and a theoretical bound developed on computational requirements.

I.2 Organization of the Dissertation

The dissertation is organized into eight chapters. This introduction is Chapter I, the remainder of which will be devoted to an overview of the contents of the subsequent chapters.

Chapter II covers the fundamental concepts of interior point methods and reviews some current research in the area. The fundamental concepts portion gives primary emphasis to the logarithmic barrier function and includes a simple example in E^2. After a brief discussion of unconstrained minimization with Newton's method, the chapter introduces the concept of computational complexity of algorithms. The chapter concludes with a survey of recent developments in analyzing the computational complexity of various interior point methods.

Chapter III begins with the rigorous presentation of the problem to be solved and the Primal-Dual Algorithm which is the core of this research effort. Several lemmas show important mathematical properties of the algorithm which will be used in later analysis.

Chapter IV analyzes the search direction in the primal variables $x \in E^n$ generated by the Primal-Dual Algorithm, and shows it to be closely related to the search direction generated by an interior point method based on the logarithmic barrier function. With careful bounding of the difference between the two directions, a primal variable algorithm using the Primal-Dual Algorithm primal variable search direction inherits convergence and polynomial complexity results of den Hertog, Roos, and Terlaky(1990) for the logarithmic barrier function method.

Chapter V follows the primal variable analysis of Chapter IV with similar analysis of the dual variable search direction. An attempt is made to show a type of monotone convergence in the dual variables toward stationarity, but examples show that such monotonicity may not occur.

Chapter VI develops a modification to the algorithm, resulting in the Modified Primal-Dual Algorithm, which adjusts the direction and/or magnitude of movement in the dual variables to produce the monotonicity sought in Chapter V. Further analysis shows that the Modified Primal-Dual Algorithm inherits the convergence and complexity results from Chapter IV.

Chapter VII presents an alternative convergence and complexity result for the

Modified Primal-Dual Algorithm based on the work of Anstreicher. By applying the Modified Primal-Dual Algorithm to a linear program with proper initialization, we show convergence and polynomial complexity precisely as in Anstreicher(1990).

Chapter VIII discusses techniques for meeting the initial feasibility and centering conditions required by the algorithm.

Chapter IX concludes the dissertation with a summary of key results and potential areas of future research.

CHAPTER II: FUNDAMENTAL CONCEPTS AND RELATED RESEARCH

II.1 Fundamental Concepts

II.1.1 Interior Point Methods

This section will present some basic results for understanding the nature and use of a barrier function as the foundation for an interior point algorithm. The material covered is not intended to be all-encompassing. Readers who desire more detail are referred to Fiacco and McCormick(1968, 1990), from which much of this discussion is drawn.

An interior point method as employed here is one type of Sequential Unconstrained Minimization Technique (SUMT). As the name implies, the idea behind SUMT is solving a series of unconstrained problems, the solutions of which converge to a solution of a constrained minimization problem. There are a variety of Sequential Unconstrained Minimization Techniques. Some allow the unconstrained problems to be solved on the boundary or outside of the feasible

region of the original problem, while others maintain strict feasibility. The later are called *interior point* methods. Again, a variety of these SUMT interior point methods exist. We will look at one of the classic types, one where a logarithmic barrier function is used to enforce feasibility, and a gradual relaxation of the barrier leads to convergence to an optimal solution.

To be more precise, let us define the general constrained minimization problem as follows:

$$\min f(x)$$
$$\text{s.t. } g_i(x) \geq 0,\ i \in \{1, 2, \ldots, m\}$$

where $x \in E^n$, and f and g_i are continuously differentiable. We make the standard assumptions of the existence of x^0 such that $g_i(x^0) > 0$, $i \in \{1, 2, \ldots, m\}$ and the existence of local minima. The unconstrained subproblem is defined in terms of the logarithmic barrier function $\Phi(x,\mu)$ for a fixed $\mu > 0$:

$$\min \Phi(x,\mu) \equiv f(x) - \mu \sum_{i=1}^{m} ln(g_i(x))$$

$$\text{s.t. } g_i(x) > 0,\ i \in \{1, 2, \ldots, m\}.$$

The logarithm of the constraint functions forces the value of $\Phi(\cdot,\mu)$ to become very large when approaching the boundary of the feasible region, so minimization of $\Phi(\cdot,\mu)$ will move away from the boundary (hence the term "barrier function"). The scalar μ is known as the barrier parameter, and its magnitude determines the

strength of this barrier. However, for any $\mu > 0$, no matter how small, $\Phi(\cdot,\mu)$ will approach $+\infty$ near the boundary of the feasible region. In practice the strict feasibility constraints are essentially ignored when minimizing $\Phi(\cdot,\mu)$, since the function's structure itself enforces feasibility. The sequence of unconstrained minimization subproblems arises from minimizing $\Phi(\cdot,\mu^k)$ for a strictly monotonically decreasing positive sequence $\{\mu^k\}$. It is shown in Fiacco and McCormick (Theorem 8) that the solutions to these subproblems will converge to a local minimizer of the constrained problem. Now suppose f(x) and $-g_i(x)$, $i \in \{1, 2, \ldots, m\}$ are convex, making the problem convex. Then local minimizers are global minimizers, and the subproblem solutions converge to a global minimizer.

The following simple example shows how such convergence occurs for a convex problem in E^2. Consider the example problem from McCormick(1983):

$$\begin{aligned} &\min \quad x_1 + x_2 \\ &\text{s.t.} \quad -x_1^2 + x_2 \geq 0. \end{aligned}$$

For $\mu > 0$ and arbitrary, the logarithmic barrier function for this problem is

$$\Phi(x,\mu) = x_1 + x_2 - \mu[ln(-x_1^2 + x_2)]$$

and its gradient and Hessian are

$$\nabla\Phi(x,\mu) = \begin{bmatrix} 1 + \dfrac{2x_1\mu}{-x_1^2 + x_2} \\ 1 - \dfrac{\mu}{-x_1^2 + x_2} \end{bmatrix}$$

$$\nabla^2\Phi(x,\mu) = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}\left(\frac{\mu}{-x_1^2 + x_2}\right)\begin{bmatrix}\sqrt{2} & 0\end{bmatrix} + \begin{bmatrix} -2x_1 \\ 1 \end{bmatrix}\left(\frac{\mu}{\left(-x_1^2 + x_2\right)^2}\right)\begin{bmatrix} -2x_1 & 1 \end{bmatrix}$$

respectively. On the interior of the feasible region, the Hessian is the sum of positive multiples of two rank one matrices. Each rank one matrix is the product of a column vector with its transpose. Therefore, the Hessian is positive semi-definite, and since the two vectors are linearly independent, the Hessian is in fact positive definite. So $\nabla\Phi(x,\mu) = 0$ within the interior of the feasible region is a sufficient condition for a strict unconstrained minimizer of $\Phi(\cdot,\mu)$. Let $x(\mu)$ denote such a minimizer. This condition implies

$$x(\mu) = \begin{bmatrix} -\frac{1}{2} & , & \frac{1}{4} + \mu \end{bmatrix}^T.$$

For instance, if $\mu^0 = 1$, then $x(\mu^0) = \begin{bmatrix} -\frac{1}{2} , & \frac{5}{4} \end{bmatrix}^T$. Solving a sequence of these subproblems for $\{\mu^k\}\downarrow 0$, we obtain the "central trajectory", and taking the limit as $\{\mu^k\}\downarrow 0$ yields $x^* = \begin{bmatrix} -\frac{1}{2} , & \frac{1}{4} \end{bmatrix}^T$, the global minimizer to the original constrained problem.

Of course, most problems do not allow an analytical solution for the location of the central trajectory. The solution of the unconstrained minimization sub-problem thus becomes a significant part of the logarithmic barrier function SUMT.

Much research has gone into unconstrained minimization in general, resulting in various candidate algorithms for minimizing $\Phi(\cdot,\mu)$ for a given μ. Steepest descent, conjugate gradient, quasi-Newton, and Newton's method are some of possible techniques. Each has advantages and disadvantages, and a detailed presentation of each is found in McCormick(1983). However, Newton's method has become the method of choice for the unconstrained sub-problem, both in the developmental days of SUMT and more recently as the speed of convergence and computational requirements of interior point algorithms have come under renewed study. The important role of Newton's method in underpinning the interior point algorithms studied in this paper calls for a brief discussion of the application of Newton's method in unconstrained minimization. This is not a rigorous mathematical development such as is found in McCormick(1983), but rather the basic ideas needed within the context of this paper.

The initial intent of Newton's method was to solve a system of n equations in n variables. The Newton search direction in this case resulted from inverting the Jacobian matrix of the system of equations and multiplying it by the negative of the system of equations. An example of this is found in Chapter III, equations (3.4) and (3.5). In the context of unconstrained minimization, the gradient and Hessian of the function to be minimized replace the system of equations and Jacobian, respectively. Thus, Newton's method is used to find a point where the gradient of a function is zero, a condition known as stationarity. Stationarity is a necessary condition for a local minimizer. If the function is strictly convex as in the previous example, this becomes a sufficient condition for a unique minimizer. We strengthen the continuity assumption to require f and $-g_i$, $i \in \{1, 2, \ldots, m\}$ to

be convex and twice continuously differentiable, and further require $\nabla^2\Phi(\cdot,\mu)$ to be positive definite on the interior of the feasible region (which implies strict convexity). From an arbitrary strictly feasible point x^k, Newton's method generates the next point in the search for stationarity as follows:

$$x^{k+1} = x^k - \left(\nabla^2\Phi(x^k,\mu)\right)^{-1}\nabla\Phi(x^k,\mu)\lambda$$

where λ is a positive step size scalar. For the "pure" Newton's method, λ is equal to unity; otherwise, the value of λ is determined by a line search along the Newton direction $-\left(\nabla^2\Phi(x^k,\mu)\right)^{-1}\nabla\Phi(x^k,\mu)$ using $\Phi(\cdot,\mu)$ as a merit function. An important way of motivating Newton's method may be seen by replacing $\Phi(\cdot,\mu)$ with a quadratic function with a positive definite Hessian. By letting $\lambda = 1$, Newton's method finds the global unconstrained minimizer in one step. Thus, when the behavior of a function between the current point in the search and the unconstrained minimizer is close to that of the quadratic approximation of the function formed at the current point, the pure Newton's method will perform very well. Indeed, as the sequence of points generated by successive pure Newton moves nears the minimizer, the quadratic approximation becomes better and better. Because of this phenomena, Newton's method with a step size of unity is said to have an asymptotic quadratic rate of convergence near the minimizer, that is, the distance from the current point to the minimizer is bounded by a constant times the square of the distance from the previous point to the minimizer. Another form of asymptotic quadratic convergence for Newtons method involves the magnitude of the Newton step. This property, which plays a key role in a number of complexity results for interior point methods, states that near the minimizer the

magnitude of the current Newton step is bounded by a constant times the square of the magnitude of the previous Newton step. However, a step size of unity may not always be possible--such a step length at times may lead to a point which violates one or more constraints and is thus not allowable.

Before moving beyond Newton's method, a few words are required about the linesearch along the Newton direction when $\lambda = 1$ is not used. Bisection or golden section search techniques are among those used. It is important to note that the strict convexity of $\Phi(\cdot, \mu)$ on the interior of the feasible region and its behavior near the boundary applies as well to $\Phi(\cdot, \mu)$ evaluated along the Newton direction, so a strict global minimizer with respect to the Newton direction exists. It will be seen later that for theoretical complexity results the type of linesearch is unimportant as long as the computational requirements for the linesearch are of the same order of magnitude as the computational requirements for generating the Newton direction.

II.1.2 Computational Complexity

The computational complexity of an algorithm is simply the theoretical worst case bound on the number of arithmetic computations required to solve a given type of problem. Such a bound is stated as a function of certain measures of problem size. Three such measures are prevalent. One is the number of bits of data required to store the problem in a computer, and is denoted as L for length. The following method for computing L for the linear programming problem $\{\min c^T x \mid Ax \geq b, x \geq 0\}$ is found in Bazaara, Jarvis, and Sherali(1990). We define

$\log(\cdot)$ as the logarithm to the base 2.

$$L = \lceil 1 + \log(n) + \log(m) + \sum_j [1 + \log(1 + |c_j|)]$$

$$+ \sum_i \sum_j [1 + \log(1 + |a_{ij}|)] + \sum_i [1 + \log(1 + |b_i|)] \rceil$$

Another measure of problem size is the dimension in Euclidian space, denoted n; and the third is m, the number of constraints. We now present the standard definition of the function $O(\cdot)$ used to express computational complexity, using notation from Papadimitriou and Steiglitz(1982). Let f(n) and h(n) denote functions from the positive integers to the positive real numbers. Then

$f(n) = O(h(n)) \Leftrightarrow \exists$ constant $k > 0$ such that for large enough n, $f(n) \leq kh(n)$.

If $f(n) = O(h(n))$, we say "f(n) is of order h(n)". One illustrative example of the nature of $O(\cdot)$ is that for any constant c, $cf(n) = O(f(n))$. An expression of computational complexity will be of the form

$$N = O(h(n))$$

where N is the total number of computations, or perhaps iterations, h is some function from the positive integers to the positive reals, and n is a positive integer or a vector of positive integers which measures the problem size. For a given algorithm, knowledge of such a function $h(\cdot)$ is key in understanding and

evaluating the complexity of the algorithm, because $h(\cdot)$ provides a bound on the growth rate of the computational requirements as the size of the problem increases.

As an illustration, we discuss the computational complexity of the simplex method for linear programming through a well-known example first shown in Klee and Minty(1972) and found in both Bazaara, Jarvis and Sherali(1990) and Papadimitriou and Steiglitz(1982). The gist of the example is that by slightly perturbing the m constraints defining a m-dimensional unit hypercube and minimizing $-x_m$ over that perturbed hypercube, the simplex method can be made to visit each vertex of the hypercube. That means the simplex method would require 2^m iterations to solve this admittedly worst-case problem, and the computational complexity of the simplex method is $O(2^m)$. Thus, increasing the dimension of the problem by one doubles the computational effort. This is an example of exponential complexity, from the exponential function K^m, $K > 1$, and indicates that the computation effort is multiplied by some number greater than one when a measure of problem size is increased by a constant, usually unity. The implication of exponential complexity is that as problem size becomes large, computational requirements grow so fast as to outstrip the capabilities of even the fastest computers. It is important to state clearly, however, that exponential complexity does not mean an algorithm is not operationally effective. The simplex method is a good example, for it has and continues to solve all manner of real world linear programming problems quickly and efficiently. The exponential complexity does indicate that there are some problems which may take a prohibitively long time to solve by the simplex method, thus prodding researchers to find another linear programming algorithm without this drawback. A more

preferable computational complexity would be expressed by a polynomial function of problem size, for instance $O(L^r m^s n^t)$ where r, s, and t are known constants. Under this type of complexity, the increase by a one of a measure of problem size results in the addition of computational work of lesser order (since, for example, $O((n+1)^t) = O(n^t)$.) Thus, polynomial complexity avoids the "exponential blow up" caused by multiplying the amount of work by some $K > 1$. The remainder of this chapter will explore recent advances in polynomial algorithms for linear programming and other classes of constrained optimization problems.

II.2 Related Research

Recent years have seen a plethora of polynomial algorithms for various classes of constrained optimization problems. This section will present an overview of several, representing some different approaches to solving the constrained optimization problem. Algorithms for linear programming are the most predominant, for at least two reasons. First, there exists the very successful and renowned simplex method, albeit with exponential complexity, so the research goal of doing "better than simplex" is tantalizing. Secondly, the linear affine constraints allow the employment of linear algebraic techniques (such as projections into the nullspace defining the feasible region) not suitable for more general nonlinear constraints. This attribute applies to quadratic programming as well, which, though not as ubiquitous as linear programming, is also heavily represented in the literature.

These algorithmic overviews are intended to provide the reader a context in which to view the complexity results presented later in this dissertation, so only a concise description of the algorithm's distinguishing characteristics is warranted. Please see the cited references for complete derivations. When an algorithm serves as an underpinning for this dissertation's results, a rigorous development of that algorithm is included at the point which it is used.

The first polynomial algorithms for constrained optimization were developed for linear programming. The first, in Khachian(1979) and the second in Karmarkar(1984) are both covered in Bazaara, Jarvis, and Sherali(1990), from which this discussion is drawn. Khachian's algorithm iteratively generates a series of shrinking ellipsoids, each of which contains the set of points solving the linear programming problem defined as minimize $\{c^T x \mid Ax \geq b,\ x \geq 0\}$. Khachian showed that the algorithm will terminate with a solution after a number of iterations bounded by a polynomial of problem size, and if that bound is exceeded no solution exists. The computational complexity bound of Khachian's algorithm is $O[(n+m)^6 L]$, where the matrix A is $m \times n$ and L is specified measure of the input length of the problem. Unfortunately, the actual computational performance of Khachian's algorithm is close to its upper bound, while the simplex method generally achieves a complexity bound of $O(m^3 n)$ in practice. Thus, Khachian's method failed in application when compared to the existing simplex method. However, the importance of Khachian's work should not be underestimated. Like the running of the first four minute mile or the breaking of the sound barrier, it showed that a feat of previously unknown feasibility could be done. This resulted in a resurgence of effort, and ultimately the "barrier" of polynomiality was broken

regularly.

While Khachian broke the barrier of polynomiality, Karmarkar introduced the first polynomial linear programming algorithm which also competed favorably in practice with the simplex method. Karmarkar's algorithm addresses a particular form of linear programming problem: minimize $\{c^T x \mid Ax = 0,\ e^T x = 1,\ x \geq 0\}$ where A is $m \times n$, A is of rank m, $n \geq 2$, data are integer, e is the vector of ones, and two fairly restrictive assumptions hold. The first assumption is that an initial point $x^0 = (1/n,\ 1/n,\ ...,\ 1/n)$ is feasible and the second is that the optimal objective function value is zero. (There are transformations and additional variables which allow standard linear programming problems to conform to these requirements.) At the heart of Karmarkar's algorithm is the following projective transformation, shown here at the start of the $(k+1)^{th}$ iteration with a feasible $x^k > 0$:

$$y_i = \frac{x_i / x_i^k}{\sum_{j=1}^{n} x_j / x_j^k}, \quad i \in \{1,\ 2,\ ...,\ n\}.$$

The algorithm takes a step in this transformed space to minimize the (transformed) objective function over the intersection of the simplex constraint $e^T y = 1$ and a ball centered at $y_0 = (1/n,\ 1/n,\ ...,\ 1/n)$ with a radius strictly less than $1/\sqrt{n(n-1)}$, the distance from y^0 to a facet of the simplex constraint. This minimization is easily accomplished by projecting the negative gradient of the transformed objective function onto the null space of the transformed feasible region and moving along it to the boundary of the constraining ball. This is the new feasible point, ready for

the next iteration. After O(nL) iterations and computational effort of $O(n^{3.5}L)$ the algorithm achieves an interior point solution x such that $c^T x \leq 2^{-L}$, at which point another polynomial algorithm may be used to find the optimal vertex. (Such an algorithm will be described later in this section.) This bound competes favorably with the $O(m^3 n)$ practical bound of the simplex method. Of particular interest to logarithmic barrier function research is the potential function used by Karmarkar to show quantifiable progress at each iteration toward optimality. That function,

$$n ln(cx) - \sum_{j=1}^{n} ln(x_j),$$

is a type of logarithmic barrier function for linear programming, and is also a component of the method of analytic centers for linear programming. This aspect of Karmarkar's algorithm spawned renewed interest in interior point methods that ultimately lead to the complexity results upon which this research is based.

Recent analysis by Powell(1991) of Karmarkar's algorithm with the potential function shown above has established that the exponent of n in the iteration bound can be no less than one. This is shown by constructing a worst-case linear program in E^3 with n inequality constraints, which transforms to $n + 3$ variables for the equality constrained problem. Karmarkar's algorithm is shown to terminate in $n/20 - 1$ iterations for $n \geq 120$ and a multiple of 20.

A well-known property of linear programming is that if a linear program has a solution, it has a vertex (or corner-point) solution. When applying any interior

point method to linear programming, it is desirable to find an optimal vertex from the nearby interior point at which the algorithm terminated. The termination criterion is that the resulting interior point $\hat{x}$ satisfies $c^T\hat{x} - z^* < 2^{-L}$, z^* being the optimal value, and is motivated by the fact that for any non-optimal vertex x, $c^Tx - z^* \notin (0\ ,\ 2^{-L})$. Since polynomiality is the stated goal of the algorithms covered, the algorithm for finding a vertex solution should also be polynomial so as not to negate the importance of that property for the linear programming algorithm. Such an algorithm is found in Bazaara, Jarvis, and Sherali(1990) and is called a purification algorithm. It is based on concepts first set forth in Charnes and Kortanek(1965). Gonzaga(1992) also discusses the basic idea of a purification algorithm: "... each iteration of a purification algorithm reduces one variable to zero along a descent direction for cost, doing pivoting like the simplex method. No more than n iterations are needed, with $O(n^2)$ computations per iteration (Gonzaga, 1992)." Gonzaga's explanation is based on an equality-constrained linear program in E^n, and it is clear that the overall complexity of this technique is of $O(n^3)$. In practice, getting to within 2^{-L} of the optimal value may be very time-consuming and therefore expensive. A practical approach, without such a stringent accuracy requirement but with no explicit consideration of polynomiality, is found in Mehrotra(1990).

Another established linear programming interior point algorithm is known as affine scaling. Gonzaga(1992) introduces affine scaling as a type of scaled-steepest descent algorithm. The nomenclature comes from the scaling of the problem at each iteration--the scaled space is achieved by dividing x_i by the current point x_i^k, $i \in \{1, 2, ..., n\}$. This maps the current point x^k to a vector of ones. The scaled

objective function negative gradient, projected onto the nullspace of the scaled constraint matrix, is the descent direction. A variety of step length procedures have been employed along the affine scaling search direction. A very simple procedure uses a constant step length of one. Another moves a fixed percentage, say 95 percent, of the distance to the nearest axis to maintain strict interiority. Because of the scaling of the feasible region, affine scaling produces an ellipsoidal confidence region in the original space, as opposed to the spherical confidence region for Cauchy's steepest descent algorithm. This allows better progress toward optimality by avoiding the "zig-zag" prevalent in unscaled steepest descent. See Gonzaga(1992) for a more thorough coverage.

Affine scaling for linear programming also exists in a primal-dual form. An example of this is found in Huang and Kortanek(1991). In this form, the search direction derives from a potential function which incorporates the primal variables (x) and the dual slacks (s). The potential function is

$$\phi(\mathrm{x},\mathrm{s}) = \rho\ ln(\mathrm{x}^{\mathrm{T}}\mathrm{s}) - \sum_{i=1}^{n} ln(\mathrm{x}_i\mathrm{s}_i)$$

where $\rho = \mathrm{n} + \sqrt{\mathrm{n}}$. The authors construct primal and dual search directions which reduce the potential function subject to primal and dual feasibility and an ellipsoidal constraint on the scaled magnitude of the search directions. The resulting algorithm simultaneously takes steps in primal and dual space at each iteration to solve a linear program with computational complexity of $O(\sqrt{n}\ L)$.

The algorithms discussed so far have been for linear programming. The

algorithm presented in this paper is for a more general class of problem with concave quadratic constraints and a convex quadratic objective function. Polynomial algorithms for solving problems more general than linear programming problems exist and are well documented, so the remainder of this chapter will present several such examples. Of course, each will apply to the more restrictive linear program as well.

A number of algorithms for solving both quadratic programs and quadratically constrained convex programs employ the method of analytic centers. General theory concerning analytic centers was developed in Huard(1967), and an example of this solution technique applied to quadratic programming is found in Mehrotra and Sun(1990). Given a general constrained optimization problem in E^n,

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in \mathbb{R} \equiv \{g_i(x) \geq 0,\ i = 1, 2, \ldots, m\}, \end{aligned}$$

for some $z \in E^1$ let the region $\mathbb{R}_z \equiv \{x \in E^n \mid f(x) \leq z,\ x \in \mathbb{R}\}$ be bounded with a non-empty interior. The analytic center of $\mathbb{R}_z$ is the point in $\mathbb{R}_z$ which solves

$$\max \ (z - f(x))^r \prod_{i=1}^{m} g_i(x).$$

for some positive integer r. An equivalent formulation was developed in Fiacco and McCormick(1968, 1990) and is related to the logarithmic barrier function:

$$\min \ -r\ln(z - f(x)) - \sum_{i=1}^{m} \ln(g_i(x)).$$

Mehrotra and Sun(1990) let r = m, a common practice. With this definition in mind, their method of analytic centers is straightforward. Let f(x) be quadratic and convex, and let each $g_i(x)$ be linear affine. We are given z^0 such that $\mathbb{R}_{z^0}$ is bounded with a non-empty interior, and x^0 a close approximation to the analytic center of $\mathbb{R}_{z^0}$. The $(k+1)^{th}$ iteration begins with z^k and x^k such that x^k a close approximation to the analytic center of $\mathbb{R}_{z^k}$. The algorithm decrements z^k to z^{k+1} and finds x^{k+1}, a close approximation to the analytic center of $\mathbb{R}_{z^{k+1}}$. The key behind the algorithm of Mehrotra and Sun is that by decrementing z^k a small amount, the algorithm requires only one partial Newton step (a step size less than one) to arrive at an approximate analytic center of $\mathbb{R}_{z^{k+1}}$. The resulting computational complexity (including the Hessian inversion for the Newton step) is $O(n^3\sqrt{m}\ L)$.

Another algorithm using analytic centers has a direct bearing on the algorithm developed later in this paper. Jarre(1991) presents a method of analytic centers to solve a convex program with quadratic constraints and a quadratic objective function, precisely the same problem solved by this dissertation's primal-dual algorithm. The iteration bound is $O(\sqrt{m}\ |ln(\epsilon)|)$ for a tolerance ϵ, also the same as will be shown for our primal-dual algorithm. The interpretation of ϵ is that the algorithm terminates with an interior point x such that the difference between the objective function evaluated at x and at the true constrained minimizer is at most ϵ. This is known as an ϵ-optimal solution. Although Jarre(1991) uses a different formulation for the definition of the analytic center than does Mehrotra and Sun(1990), the concept is the same. Given a small decrement of z^k, the analysis

shows that a single Newton step will arrive at a new approximate analytic center. Jarre also briefly discusses the problem of finding the initial values of z^0 and x^0. and shows that by adding an artificial variable and constraint the same polynomial method of analytic centers can be used to find z^0 and x^0. See Chapter VIII for an explanation and adaptation of this technique for meeting the initial condition requirements of our primal-dual algorithm.

The emergence of the user-defined parameter ϵ warrants some discussion. Because it is user-defined, ϵ is not related to problem size. A polynomial complexity bound containing ϵ, such as that of Jarre(1991) or that of our primal-dual algorithm, may thus seem less satisfying than those for linear programming which contain only problem size information, namely n, m, and L. The requirement for ϵ is a result of the more general quadratic objective function and constraints. Assume a solution exists. With a nonlinear convex quadratic objective function, an extreme point solution is not guaranteed; and even with a linear objective function, thought an extreme point solution must exist, it will not necessarily be at a "vertex". Indeed, the concept of a vertex may not apply in the presence of nonlinear quadratic constraints. For example, the concave quadratic constraint $g(x) = 1 - x^2 - y^2 \geq 0$ in E^2 generates a feasible region consisting of the boundary and interior of the unit sphere, which is compact but clearly has no vertices. Thus, there is no way for some measure of problem size to indicate a stopping point near an optimal vertex as L does for linear programming. Hence, the use of a tolerance parameter ϵ. One must take care, however, that the manner in which ϵ appears in the complexity bound does not compromise polynomiality. As a reader of this dissertation pointed out, had a bound of $O(\sqrt{m}\, 2^{|ln\epsilon|})$ resulted,

substitution of 2^{-L} for ϵ as is done for linear programming would result in exponential complexity even though $2^{|ln\epsilon|}$ is a user-defined scalar. The scale of the problem also may complicate the selection of ϵ. For an arbitrary $\epsilon > 0$ and objective function f(x), suppose we can compute T such that $|f(x_j)| \leq T$ for all feasible x. Then with the simple transformation $\hat{f}(x) = (\epsilon/2M)f(x)$, minimizing $\hat{f}(x)$ over the feasible region results in every feasible x being an ϵ-optimal solution. So while the use of the tolerance ϵ is necessary for the result, readers should recognize that polynomiality based partly on ϵ does differ from polynomiality based strictly on problem size.

The research which provides the underpinnings to this dissertation concerns the logarithmic barrier function and SUMT for solving both the convex quadratically constrained problem and the quadratic program. Significant research in the recent past has proven polynomiality for this solution technique for both of these problem classes. Two papers in particular are crucial to our results--they are mentioned briefly here, and covered in detail later in subsequent chapters as they apply. den Hertog, Roos, and Terlaky(1990) showed that the classic logarithmic barrier method solves the quadratically constrained problem to within a tolerance of ϵ in $O(\sqrt{m}\ |ln(\epsilon)|)$ iterations. The convergence and complexity analysis of den Hertog, Roos, and Terlaky provides the basis for the initial complexity result for our primal-dual algorithm, and is presented in detail in Chapter IV. Anstreicher(1990) applied the logarithmic barrier method to the quadratic program. The convergence and complexity analysis differs substantially from den Hertog, Roos, and Terlaky(1990), and establishes an iteration bound of $O(\sqrt{m}\ L)$. Furthermore, Anstreicher shows that with the correct initialization and selection of

parameters, the classic logarithmic barrier SUMT as implemented by Fiacco and McCormick in the 1960s has polynomial complexity of $O(\sqrt{m}\ L\ ln(L))$ for quadratic programming. Anstreicher's results are presented in Chapter VII, where they establish that the primal-dual algorithm developed in this dissertation has an $O(\sqrt{m}\ L)$ iteration bound when applied to a linear programming problem. (This particular bound does not extend to quadratic programming.)

This chapter would be incomplete without citing two recent and extensive survey papers on interior point methods. The first, Gonzaga(1992), has already been cited with respect to an affine scaling linear programming algorithm. It contains a wealth of information on primal, dual, and primal-dual interior point methods for linear programming. The second is Wright(1992), which is an excellent primer on the basic theory underlying interior point methods for linear and nonlinear programming, with special emphasis on the logarithmic barrier function. Wright includes a polynomial complexity proof for a linear programing barrier function method that is somewhat similar to Anstreicher's, as well as a survey of primal and primal-dual methods and linear algebraic issues for interior point methods.

Finally, in almost any reference concerning barrier functions in mathematical programming, including those above, one will find citations for Fiacco and McCormick(1968, 1990). Their classic text on nonlinear programming, first published in 1968 and now reissued, remains a fundamental source for understanding this field of research. Their efforts at the time were directed toward nonlinear programming and did not involve computational complexity, but did lay

the groundwork for much of what has followed in linear and nonlinear programming. The startling and satisfying discovery by Anstreicher of the polynomiality of their original 1960s-vintage SUMT for quadratic programming indicates the enduring value of their research.

CHAPTER III: THE PRIMAL-DUAL ALGORITHM

III.1 Problem Definition

We define a convex problem P:

$$\begin{aligned} &\min \ b^{T}x \\ &\text{s.t. } g_i(x) \geq 0, \ i \in I \equiv \{1,2, \ldots ,m\} \\ &\quad x \in E^n \end{aligned}$$

where b is a vector in E^n and each $g_i(x)$ is a concave quadratic function for $i \in I$. Let $\mathbb{R} \equiv \{x \in E^n \mid g_i(x) \geq 0, i \in I\}$ and $\mathbb{R}^* \equiv \{x \in E^n \mid g_i(x) > 0, i \in I\}$. Let us assume $\mathbb{R}$ is bounded and $\mathbb{R}^*$ is non-empty; and let us define the following logarithmic barrier function on $\mathbb{R}^*$, for $\mu \in E^1$ and strictly positive:

$$\Phi(x,\mu) \equiv \frac{b^{T}x}{\mu} - \sum_{i=1}^{m} \ln(g_i(x)).$$

The gradient and Hessian of this function are as follows:

$$\nabla\Phi(x,\mu) = \frac{b}{\mu} - \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)}$$

$$\nabla^2\Phi(x,\mu) = \sum_{i=1}^{m}\left[\frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}\right]$$

Note that $\Phi(x,\mu)$ is convex in $\mathbb{R}^*$ for fixed $\mu > 0$ and thus has a positive semi-definite Hessian. This structure allows for the following lemmas:

Lemma III.1: For an arbitrary $z \in E^n$ and nonzero, and $x \in \mathbb{R}^*$,

If $\mathbb{R}$ is bounded, then

$$z^T\nabla g_i(x)\nabla g_i(x)^T z = 0,\ i \in I \tag{3.1}$$

and

$$z^T\nabla^2 g_i(x)\ z = 0,\ i \in I \tag{3.2}$$

cannot both be true.

Proof: Suppose not. Then (3.1) and (3.2) both hold. Let $i \in I$ be arbitrary. That (3.1) holds implies $\nabla g_i(x)^T z = 0$. Now let $y = x + \lambda z$, $\lambda \in E^1$. Using $x \in \mathbb{R}^*$ and (3.2) leads to

$$g_i(y) = g_i(x) + \lambda\nabla g_i(x)^T z + \frac{\lambda^2}{2} z^T\nabla^2 g_i(x)\ z = g_i(x) > 0.$$

Since i was arbitrary, this holds for $i \in I$, which implies $y \in \mathbb{R}^*$. But $\| y \| \to +\infty$ as

$\lambda \to +\infty$, so $\mathbb{R}^*$ is unbounded. This contradiction shows the supposition to be false, proving the lemma. □

Lemma III.2: If $\mathbb{R}$ is bounded, then $\nabla_x^2\Phi(x,\mu)$ is positive definite in $\mathbb{R}^*$.

Proof: Let $z \in E^n$ and nonzero be given.

$$z^T\nabla_x^2\Phi(x,\mu)z = z^T\left[\sum_{i=1}^{m}\left(\frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}\right)\right]z \tag{3.3}$$

$$= \sum_{i=1}^{m}\left(\frac{z^T\nabla g_i(x)\nabla g_i(x)^T z}{g_i(x)^2}\right) + \sum_{i=1}^{m}\left(\frac{z^T(-\nabla^2 g_i(x))z}{g_i(x)}\right)$$

Since $g_i(x) > 0$, and $\nabla g_i(x)\nabla g_i(x)^T$ and $(-\nabla^2 g_i(x))$ are positive semi-definite, $i \in I$, each term in each summation is non-negative. Then by Lemma III.1, at least one term must be strictly positive, proving the lemma. □

The specification of a linear objective function in problem P is not restrictive. Any constrained optimization problem with a general objective function r(x) to be minimized can be transformed into a constrained optimization problem with a linear objective function by defining another variable z, adding the constraint $z - r(x) \geq 0$, and minimizing z subject to this and the original constraints. Thus, given a problem

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } g_i(x) \geq 0,\ i \in I \\ &\quad x \in E^n \end{aligned}$$

where each g_i is a concave quadratic and f is a non-linear convex quadratic function, one may convert it to the form of problem P by introducing an additional variable $z \in E^1$ and using the transformation above to achieve

$$
\begin{aligned}
P_z\text{: } & \min z \\
& \text{s.t. } g_i(x) \geq 0,\ i \in I \\
& \quad z - f(x) \geq 0 \\
& \quad x \in E^n,\ z \in E^1.
\end{aligned}
$$

This is an equivalent problem but with a linear objective function and concave quadratic constraints. Note the implication that a convex quadratic program (QP) is therefore a special case of P.

This transformation does not compromise the ϵ-optimality criterion mentioned in section II.2 and in Jarre(1991), as we now show. Let (x^*, z^*) solve P_z, and let (x^k, z^k) be an ϵ-optimal solution for P_z. By construction, x^* solves the original problem and $z^* = f(x^*)$. Since (x^k, z^k) is feasible we have $f(x^k) \leq z^k$ and $f(x^k) - z^k \leq 0$. Then using the definition of ϵ-optimality,

$$
\begin{aligned}
& z^k - z^* \leq \epsilon \\
& z^k + f(x^k) - z^k - z^* \leq \epsilon \\
& f(x^k) - z^* \leq \epsilon \\
& f(x^k) - f(x^*) \leq \epsilon
\end{aligned}
$$

So x^k is an ϵ-optimal solution to the original problem.

Since P is a convex problem, there exist first order sufficiency conditions for a global minimizer for P. We first define the Lagrangian for P using the notation $u \equiv (u_1, u_2, \ldots, u_m)$:

$$L(x,u) \equiv b^T x - \sum_{i=1}^{m} u_i g_i(x)$$

Henceforth, we will let $\nabla L(x,u)$ denote $\nabla_x L(x,u)$.

First Order Sufficiency Conditions for P (see McCormick(1983) and Fiacco and McCormick(1968, 1990))

$$\exists\, \bar{x} \in \mathbb{R},\ \bar{u} \geq 0 \text{ such that } \nabla L(\bar{x}, \bar{u}) = 0 \text{ and } \bar{u}_i g_i(\bar{x}) = 0,\ i \in I$$

$$\Rightarrow \bar{x} \text{ is a global minimizer for P}$$

We also have the associated Wolfe dual. The assumption that $\mathbb{R}^*$ is non-empty satisfies Slater's condition for convex programming, so $(\bar{x}, \bar{u})$ solves the dual problem D (see Fiacco and McCormick(1968, 1990)):

$$D: \max\ L(x,u) = b^T x - \sum_{i=1}^{m} u_i g_i(x)$$

$$\text{s.t. } \nabla L(x,u) = b - \sum_{i=1}^{m} u_i \nabla g_i(x) = 0,\ u_i \geq 0,\ i \in I$$

Returning to the logarithmic barrier function as defined for problem P, consider the problem

$$\begin{aligned} &\min \Phi(x,\mu) \\ &\text{s.t. } x \in \mathbb{R}^* \end{aligned}$$

for an arbitrary $\mu > 0$. Let $x(\mu)$ be the minimizer of $\Phi(x,\mu)$ in $\mathbb{R}^*$. By defining $u_i(\mu) \equiv \mu / g_i(x(\mu))$ and $u(\mu) \equiv \{u_1(\mu), u_2(\mu), \ldots, u_m(\mu)\}$ we see that $(x(\mu), u(\mu))$ is dual feasible. This dual feasibility allows the formulation of a lower bound on the optimal value of the objective function $b^T\overline{x}$ (see Fiacco and McCormick(1968, 1990), Theorem 22). This lower bound combined with the upper bound given by $b^Tx(\mu)$ (since $x(\mu)$ is feasible) yields

$$b^Tx(\mu) - \sum_{i=1}^{m} u_i(\mu) g_i(x(\mu)) \leq b^T\overline{x} \leq b^Tx(\mu)$$

$$\Rightarrow \quad b^Tx(\mu) - m\mu \leq b^T\overline{x} \leq b^Tx(\mu).$$

A more extended development of duality in the minimization of $\Phi(x, \mu)$ is found in Fiacco and McCormick(1968, 1990), the text upon which much of the preceding discussion is based. An important lemma relating to these upper and lower bounds when minimizing $\Phi(x, \mu)$ is contained in den Hertog, Roos, and Terlaky(1990), and is presented here without proof.

Lemma III.3: For decreasing values of $\mu > 0$, the objective function $b^Tx(\mu)$ is strictly monotonically decreasing and the dual objective function $b^Tx(\mu) - m\mu$ is

monotonically increasing.

Proof: Omitted. See den Hertog, Roos and Terlaky(1990), where the authors prove the result for general smooth convex programs. The first part was first proved in Fiacco and McCormick(1968, 1990) for the more general convex programming problem. □

III.2 The Primal-Dual Algorithm

This research concerns convergence and computational complexity of an algorithm that, applied to problem P, simultaneously finds the minimizer $\bar{x}$ and the associated vector of dual variables (i.e. KKT multipliers) $\bar{u}$. The following motivation for the algorithm is from McCormick(1991b).

Primal-Dual Algorithm: Solve the equations

$$\nabla L(x,u) = 0 \tag{3.4}$$

$$u_i g_i(x) = \mu^k, \; i \in I \tag{3.5}$$

with $x \in \mathbb{R}^*$ and $u_i > 0$, $i \in I$, for a positive sequence $\{\mu^k\} \downarrow 0$. Satisfaction of (3.5) is known as perturbed complementary slackness. For a given value of $\mu^k > 0$, let $[x(\mu^k), u(\mu^k)]$ denote a point satisfying the above equations with $x(\mu^k) \in \mathbb{R}^*$ and $u_i(\mu^k) > 0$, $i \in I$. Then $[x(\mu^k), u(\mu^k)]$ is dual feasible, and

$$b^T x(\mu^k) \geq b^T \bar{x} \geq b^T x(\mu^k) - \sum_{i=1}^{m} u_i(\mu^k) g_i(x(\mu^k)) = b^T x(\mu^k) - m\mu^k.$$

The point $x(\mu^k)$ is also a minimizer of $\Phi(x,\mu^k)$; this can be seen by substituting $\mu^k / g_i(x(\mu^k))$ for $u_i(\mu^k)$ in $\nabla L(x(\mu^k), u(\mu^k)) = 0$, which results in $\nabla\Phi(x(\mu^k),\mu^k) = 0$. Φ inherits the convexity of problem P, so stationarity and feasibility are sufficient for a global minimizer.

In practice, these equations would not be solved exactly, but approximately. The basic idea of the algorithm is to use Newton's method to solve (3.4) and (3.5) with a line search to determine step size and retain feasibility. We assume that at the start of the k^{th} Newton iteration we have $\mu^k > 0$ and (x,u) such that $x \in \mathbb{R}^*$ and $u_i > 0$. $i \in I$. Let $U \equiv \text{diag}(u_i)$, $G \equiv \text{diag}(g_i(x))$, and ∇g be the $n \times m$ matrix whose i^{th} column is $\nabla g_i(x)$. For simpler notation, let μ denote μ^k. The Jacobian of (3.4) and (3.5) is

$$J(x,u) = \begin{bmatrix} \nabla^2 L(x,u) & -\nabla g \\ U\nabla g^T & G \end{bmatrix}$$

Using the bordered inverse formula in McCormick(1983), we have

$$J^{-1}(x,u) = \begin{bmatrix} M^{-1} & M^{-1}\nabla g\, G^{-1} \\ -G^{-1}U\nabla g^T M^{-1} & D \end{bmatrix}$$

where $D = G^{-1} - G^{-1}U\nabla g^T M^{-1}\nabla g\, G^{-1}$, and M is defined as follows:

$$M \equiv \nabla^2 L(x,u) + \nabla g\, UG^{-1}\nabla g^T$$

$$= \nabla^2 L(x,u) + \sum_{i=1}^{m} \nabla g_i(x) \frac{u_i}{g_i(x)} \nabla g_i(x)^T$$

for $x \in R^*$ and $u_i > 0$, $i \in I$. Some manipulation reveals

$$M = \sum_{i=1}^{m} u_i g_i(x) \left[\frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right]. \qquad (3.6)$$

The following lemma results from the structure of M and its similarity to the Hessian of $\Phi(x,\mu)$.

Lemma III.4: R is bounded $\Rightarrow$ M is positive definite in $\mathbb{R}^*$, for $u_i > 0$, $i \in I$.

Proof: Let $z \in E^n$ and nonzero be given.

$$z^T M z = z^T \left[\sum_{i=1}^{m} u_i g_i(x) \left[\frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right] \right] z$$

$$= \sum_{i=1}^{m} \left(\frac{u_i z^T \nabla g_i(x) \nabla g_i(x)^T z}{g_i(x)} \right) + \sum_{i=1}^{m} \left(u_i z^T (- \nabla^2 g_i(x)) z \right)$$

Since $u_i > 0$, $g_i(x) > 0$, and $\nabla g_i(x) \nabla g_i(x)^T$ and $(- \nabla^2 g_i(x))$ are positive semi-definite, $i \in I$, each term in each summation is non-negative. Then by Lemma III.1, at least one term must be strictly positive, proving the lemma. □

Note that for positive definite M and $\mu > 0$,

$$\frac{1}{\mu}\mathrm{M} = \sum_{i=1}^{m} \frac{\mathrm{u}_i g_i(\mathrm{x})}{\mu}\left[\frac{\nabla g_i(\mathrm{x})\nabla g_i(\mathrm{x})^{\mathrm{T}}}{g_i(\mathrm{x})^2} - \frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\right] \tag{3.7}$$

$$= \sum_{i=1}^{m}\left[\frac{\nabla g_i(\mathrm{x})\nabla g_i(\mathrm{x})^{\mathrm{T}}}{g_i(\mathrm{x})^2} - \frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\right]$$

$$- \sum_{i=1}^{m}\left(1 - \frac{\mathrm{u}_i g_i(\mathrm{x})}{\mu}\right)\left[\frac{\nabla g_i(\mathrm{x})\nabla g_i(\mathrm{x})^{\mathrm{T}}}{g_i(\mathrm{x})^2} - \frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\right]$$

$$= \nabla_x^2 \Phi(\mathrm{x},\mu) - \sum_{i=1}^{m}\left(1 - \frac{\mathrm{u}_i g_i(\mathrm{x})}{\mu}\right)\left[\frac{\nabla g_i(\mathrm{x})\nabla g_i(\mathrm{x})^{\mathrm{T}}}{g_i(\mathrm{x})^2} - \frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\right] \tag{3.8}$$

and is positive definite.

We have shown that for bounded $\mathbb{R}$, M^{-1} exists. If $\mathrm{x} \in \mathbb{R}^*$ then G^{-1} exists. So given the boundedness of $\mathbb{R}$ and interiority of x, $\mathrm{J}^{-1}(\mathrm{x},\mathrm{u})$ exists. Note that the diagonal nature of G and U implies that the effort in computing $\mathrm{J}^{-1}(\mathrm{x},\mathrm{u})$ is of the same order as in computing M^{-1}, that is, $O(\mathrm{n}^3)$.

This explicit expression for the Jacobian inverse allows the formulation of the Newton step with step size t. Here, e is a vector of ones.

$$\begin{bmatrix} \mathrm{x}(\mathrm{t}) \\ \mathrm{u}(\mathrm{t}) \end{bmatrix} = \begin{bmatrix} \mathrm{x} \\ \mathrm{u} \end{bmatrix} - \begin{bmatrix} \mathrm{M}^{-1} & \mathrm{M}^{-1}\nabla g\ \mathrm{G}^{-1} \\ -\mathrm{G}^{-1}\mathrm{U}\nabla g^{\mathrm{T}}\mathrm{M}^{-1} & \mathrm{D} \end{bmatrix}\begin{bmatrix} \nabla \mathrm{L}(\mathrm{x},\mathrm{u}) \\ \mathrm{UG}e - \epsilon\mu \end{bmatrix} \mathrm{t}$$

$$= \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} -M^{-1}\left[b^T - \mu\sum_{i=1}^{m} \nabla g_i(x)/g_i(x)\right] \\ UG^{-1}\nabla g^T M^{-1}\left[b^T - \mu\sum_{i=1}^{m} \nabla g_i(x)/g_i(x)\right] + G^{-1}e\mu - u \end{bmatrix} t$$

Now let $p \equiv -M^{-1}\left[b^T - \mu\sum_{i=1}^{m} \nabla g_i(x)/g_i(x)\right] = -\left(\frac{1}{\mu}M\right)^{-1}\nabla\Phi(x,\mu)$.

The Newton step becomes

$$\begin{bmatrix} x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} p \\ -UG^{-1}\nabla g^T p + G^{-1}e\mu - u \end{bmatrix} t$$

We next explore the implications of the Primal-Dual Newton direction in the primal variables.

CHAPTER IV: PROPERTIES OF THE PRIMAL-DUAL NEWTON DIRECTION IN THE PRIMAL VARIABLES

IV.1 Developing a Primal Variable Algorithm

Let H denote $\nabla^2\Phi(x,\mu)$. We have noted the strong connection between H and M/μ for fixed $\mu > 0$, namely.

$$\frac{1}{\mu}M = H - \sum_{i=1}^{m}\left(1 - \frac{u_i g_i(x)}{\mu}\right)\left[\frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}\right].$$

Indeed, if $u_i g_i(x) - \mu = 0$ for all $i \in I$, then $M/\mu = H$.

Now let $\omega = \max_{i \in I}\left|1 - \frac{u_i g_i(x)}{\mu}\right|$. Then for any $d \in E^n$,

$$d^T\left[H - \omega\sum_{i=1}^{m}\left|\frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}\right|\right]d \leq d^T\frac{1}{\mu}M\,d$$

$$\leq \mathrm{d}^{\mathrm{T}}\left[\mathrm{H}+\omega\sum_{i=1}^{m}\left[\frac{\nabla g_i(\mathrm{x})\nabla g_i(\mathrm{x})^{\mathrm{T}}}{g_i(\mathrm{x})^2}-\frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\right]\right]\mathrm{d}$$

which implies

$$(1-\omega)\mathrm{d}^{\mathrm{T}}\mathrm{H}\,\mathrm{d} \;\leq\; \mathrm{d}^{\mathrm{T}}\tfrac{1}{\mu}\mathrm{M}\,\mathrm{d} \;\leq\; (1+\omega)\mathrm{d}^{\mathrm{T}}\mathrm{H}\,\mathrm{d}, \tag{4.1}$$

and if $\omega < 1$,

$$\sqrt{1-\omega}\,\|\mathrm{d}\|_H \;\leq\; \|\mathrm{d}\|_{M/\mu} \;\leq\; \sqrt{1+\omega}\,\|\mathrm{d}\|_H \tag{4.2}$$

where for any positive definite $\mathrm{n}\times\mathrm{n}$ matrix A, $\|\mathrm{d}\|_A \equiv (\mathrm{d}^{\mathrm{T}}\mathrm{A}\,\mathrm{d})^{1/2}$ defines the matrix norm of d with respect to A. We will now adapt the results of den Hertog, Roos, and Terlaky(1990) to develop a primal variable algorithm using the primal variable Newton direction $\mathrm{p}(\mathrm{x},\mathrm{u},\mu) = -(\mathrm{M}/\mu)^{-1}\nabla\Phi(\mathrm{x},\mu)$ generated by the Primal-Dual Algorithm, and ω sufficiently small to obtain convergence and polynomial complexity. We will use p as an abbreviation for $\mathrm{p}(\mathrm{x},\mathrm{u},\mu)$ when clarity does not suffer because of it.

A Primal Variable Algorithm: This algorithm loosely follows the SUMT central path. It gets close to the central path by doing linesearches seeking to reduce $\Phi(\mathrm{x},\mu)$ along successive Newton directions. When used in this manner in the course of a line search, $\Phi(\mathrm{x},\mu)$ is called a merit function. The Newton direction p is the one generated in the primal variables by the Primal-Dual Algorithm. We

assume

$$\omega = \max_{i \in I} \left| 1 - \frac{u_i g_i(x)}{\mu} \right| \leq \frac{1}{62}$$

at all times. This assumption is not blithely made--it is quite restrictive, and the reader may reasonably wonder whether it could actually be satisfied without sacrificing polynomiality. In fact, Chapter VI of this dissertation presents a method for doing so.

It will be shown that this bound on ω keeps M/μ sufficiently close to H to allow proofs of convergence and polynomial complexity. The algorithm proceeds by doing successive linesearches until $2\omega\sqrt{1+\omega} \leq \|p\|_{M/\mu} \leq \tau = \frac{1}{8}\sqrt{1-\omega}$, a tolerance. The positive lower bound on this tolerance is quite small because of the size of ω, so if it is approached during the linesearch the upper bound will be easily met, and the linesearch may be terminated. (For $\omega \leq \frac{1}{16}$, the upper bound is strictly greater than the lower bound.) When this tolerance is obtained, we reduce μ to $\overline{\mu} = (1-\theta)\mu$ for some $\theta \in (0,1)$. The algorithm ultimately achieves an ϵ-optimal solution value for P. A point x is said to be ϵ-optimal if, given the true constrained minimum z^*, $b^T x - z^* \leq \epsilon$.

Input

Reduction factor $\theta \in (0,1)$

Tolerance $\tau = \frac{1}{8}\sqrt{1-\omega}$. Assume $\omega = \max_{i \in I} \left| 1 - \frac{u_i g_i(x)}{\mu} \right| \leq \frac{1}{62}$ always.

Υ is an accuracy parameter.

$x^0 \in \mathbb{R}^*$ is given, as is $\mu^0 \leq 1/\Upsilon$, and $u_i^0 > 0$, $i \in I$ such that $\left\| p(x^0,u^0,\mu^0) \right\|_{M/\mu} \leq \tau$

begin

$x = x^0$, $u = u^0$, $\mu = \mu^0$

while $\mu > \Upsilon/2m$ do

begin outer iteration

while $\| p \|_{M/\mu} \leq \tau$ do

begin inner iteration

$\overline{\lambda} = \arg \min_{0 < \lambda \leq 1} \{\Phi(x,\mu)$ s.t. $(x + \lambda p) \in \mathbb{R}^*$, $\| p \|_{M/\mu} \geq 2\omega\sqrt{1+\omega}\}$

$x = x + \overline{\lambda} p$

end inner iteration

$\mu = (1-\theta)\mu$

end outer iteration

end

IV.2 Convergence and Complexity of the Primal Variable Algorithm

Let $q_x(x + d, \mu) \equiv \Phi(x,\mu) + \nabla\Phi(x,\mu)^T d + \frac{1}{2} d^T H d$, the quadratic approximation of $\Phi(x + d, \mu)$ at x.

Let $\hat{q}_x(x + d, \mu) \equiv \Phi(x,\mu) + \nabla\Phi(x,\mu)^T d + \frac{1}{2} d^T \frac{1}{\mu} M d$

$$= \Phi(x,\mu) + \nabla\Phi(x,\mu)^T d + \tfrac{1}{2}\, d^T(H - E)\, d$$

$$= q_x(x + d,\, \mu) - \tfrac{1}{2}\, d^T E\, d$$

where $E \equiv \sum_{i=1}^{m}\left(1 - \frac{u_i g_i(x)}{\mu}\right)\left[\frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}\right]$

From the relationship between H and $\frac{1}{\mu}M$ and the definition of ω we have

$$q_x(x + d,\, \mu) - \tfrac{1}{2}\, \omega d^T H\, d \;\le\; \hat{q}_x(x + d,\, \mu) \;\le\; q_x(x + d,\, \mu) + \tfrac{1}{2}\, \omega d^T H\, d$$

The proof of convergence and polynomial complexity for the primal variable algorithm is developed through a series of five lemmas. The synopsis below is based on one in den Hertog, Roos, and Terlaky(1990), and may aid the reader in following the development of the convergence and complexity proof:

(1) Lemma IV.1 gives an upper bound on the absolute difference between $\Phi(x + d,\mu)$ and $\hat{q}_x(x + d,\mu)$.

(2) Lemma IV.2 shows that if a proximity criterion based on the Primal-Dual Newton step in the primal variables is met at a point x, then x lies close to the minimizer of $\Phi(\cdot\,,\mu)$. "Closeness" is measured using the norm with respect to the Hessian of $\Phi(x,\mu)$.

(3) Lemma IV.3 establishes that when the proximity criterion from Lemma

IV.2 is not met, one obtains a guaranteed reduction in $\Phi(\cdot,\mu)$ when doing a line search along the Primal-Dual Newton step in the primal variables.

(4) Lemma IV.4 gives an upper bound on the difference between $\Phi(x,\mu)$ and $\Phi(x(\mu),\mu)$, where $x(\mu)$ is the minimizer of $\Phi(\cdot,\mu)$, when the proximity criterion from Lemma IV.2 is met.

(5) Lemma IV.5 gives a bound on the absolute difference between the values obtained by evaluating the objective function of P at x and $x(\mu)$, when the proximity criterion from Lemma IV.2 is met.

With our previous assumption of $\omega \leq \frac{1}{62}$ throughout all iterations, we can prove these lemmas analogous to those in den Hertog, Roos, and Terlaky(1990). The proofs are similar to those in the referenced work, adapted to deal with the use of M/μ instead of H in the quadratic approximation of $\Phi(x+d,\mu)$.

Lemma IV.1: Let $\omega \leq \frac{1}{62}$, $\|d\|_{M/\mu} < \sqrt{1-\omega}$ for an arbitrary $d \in E^n$, and let $x \in \mathbb{R}^*$, $\mu > 0$ be given.
Then $x+d \in \mathbb{R}^*$ and

$$\left|\Phi(x+d,\mu) - \hat{q}_x(x+d,\mu)\right| \leq \tfrac{1}{2}\omega\|d\|_H^2 + \frac{\|d\|_H^3}{3(1-\|d\|_H)}$$

Proof: $\|d\|_{M/\mu} < \sqrt{1-\omega} \Rightarrow \|d\|_H < 1$. Using the Taylor expansion of $\Phi(x,\mu)$, we have

$$\Phi(x+d,\,\mu) = q_x(x+d,\,\mu) + \sum_{i=3}^{\infty} t_i$$

where t_i is the i^{th} term of the Taylor expansion. In den Hertog, Roos, and Terlaky(1990) it is proved that for $\|d\|_H < 1$, $x + d \in \mathbb{R}^*$ and $\sum_{i=3}^{\infty} t_i$ converges such that

$$\left| \sum_{i=3}^{\infty} t_i \right| \leq \frac{\|d\|_H^3}{3(1-\|d\|_H)}$$

Thus, $$\Phi(x+d,\,\mu) = \hat{q}_x(x+d,\,\mu) + \tfrac{1}{2}\, d^T E\, d + \sum_{i=3}^{\infty} t_i$$

$$\Phi(x+d,\,\mu) - \hat{q}_x(x+d,\,\mu) \;\leq\; \tfrac{1}{2}\, d^T \omega H\, d + \sum_{i=3}^{\infty} t_i$$

$$\left|\Phi(x+d,\,\mu) - \hat{q}_x(x+d,\,\mu)\right| \;\leq\; \tfrac{1}{2}\,\omega \|d\|_H^2 + \frac{\|d\|_H^3}{3(1-\|d\|_H)},$$ proving the lemma. □

Lemma IV.2: If $\omega \leq \frac{1}{62}$ and $2\omega\sqrt{1+\omega} \leq \|p\|_{M/\mu} \leq \frac{1}{8}\sqrt{1-\omega}$, then $\|x - x(\mu)\|_H < \frac{5}{2}\|p\|_H$

Proof: $\|p\|_{M/\mu} < \frac{1}{8}\sqrt{1-\omega} \;\Rightarrow\; \|p\|_H \leq \frac{1}{8}$. Let h be arbitrary such that $\|h\|_H = \frac{3}{2}\|p\|_H$.

$$\|p+h\|_H \;\leq\; \|p\|_H + \|h\|_H \;=\; \tfrac{5}{2}\|p\|_H \;<\; \tfrac{1}{3}. \qquad (4.3)$$

Consider values of Φ evaluated on the ellipsoid $\{x+p+h \mid \|h\|_H = \frac{3}{2}\|p\|_H\}$. From Lemma IV.1,

$$\Phi(x+p+h,\,\mu) \;\geq\; \hat{q}_x(x+p+h,\,\mu) \;-\; \tfrac{1}{2}\,\omega\|p+h\|_H^2 - \frac{\|p+h\|_H^3}{3(1-\|p+h\|_H)} \tag{4.4}$$

Since $x+p$ minimizes $\hat{q}_x$, $\nabla\hat{q}_x(x+p,\,\mu)=0 \;\Rightarrow$

$$\hat{q}_x(x+p+h,\,\mu) = \hat{q}_x(x+p,\,\mu) + \nabla\hat{q}_x(x+p,\,\mu)^T h + \tfrac{1}{2}h^T\nabla^2\hat{q}_x(x+p,\,\mu)h$$

$$= \hat{q}_x(x+p,\,\mu) + \tfrac{1}{2}\,h^T(\tfrac{1}{\mu}M)\,h \tag{4.5}$$

Substituting (4.5) and (4.3) into (4.4) yields

$$\Phi(x+p+h,\,\mu) \;\geq\; \hat{q}_x(x+p,\,\mu) + \tfrac{1}{2}\,h^T(\tfrac{1}{\mu}M)\,h \;-\; \tfrac{1}{2}\,\omega\tfrac{25}{4}\,\|p\|_H^2 - \frac{\frac{125}{8}\|p\|_H^3}{3(1-\frac{1}{3})}$$

$$\geq\; \hat{q}_x(x+p,\,\mu) + \tfrac{1}{2}\,(1-\omega)\|h\|_H^2 \;-\; \omega\tfrac{25}{8}\,\|p\|_H^2 - \tfrac{125}{16}\|p\|_H^3$$

$$\geq\; \hat{q}_x(x+p,\,\mu) + \tfrac{1}{2}\,(1-\omega)\tfrac{9}{4}\,\|p\|_H^2 \;-\; \omega\tfrac{25}{8}\,\|p\|_H^2 - \tfrac{125}{16}\|p\|_H^3$$

$$\geq\; \hat{q}_x(x+p,\,\mu) + \left(\tfrac{9}{8} - \tfrac{34}{8}\omega\right)\|p\|_H^2 \;-\tfrac{125}{16}\|p\|_H^3$$

$\omega \leq \frac{1}{62} \;\Rightarrow 9-34\omega \geq 0$, which with $\|p\|_H \leq \frac{1}{8}$ implies

$$\Phi(x+p+h,\,\mu) \;\geq\; \hat{q}_x(x+p,\,\mu) + (9\;-34\omega)\,\|p\|_H^3 \;-\tfrac{125}{16}\|p\|_H^3$$

$$\geq \widehat{q}_x(x+p,\mu) + \left(\tfrac{19}{16} - 34\omega\right)\|p\|_H^3$$

From Lemma IV.1 we also have

$$\Phi(x+p,\mu) < \widehat{q}_x(x+p,\mu) + \tfrac{1}{2}\omega\|p\|_H^2 + \frac{\|p\|_H^3}{3(1-\|p\|_H)}$$

Since $2\omega\sqrt{1+\omega} \leq \|p\|_{M/\mu} \Rightarrow \omega \leq \frac{1}{2}\|p\|_H$, we have

$$\Phi(x+p,\mu) < \widehat{q}_x(x+p,\mu) + \tfrac{1}{4}\|p\|_H^3 + \frac{\|p\|_H^3}{3(1-\frac{1}{8})}$$

$$< \widehat{q}_x(x+p,\mu) + \tfrac{53}{84}\|p\|_H^3$$

and so $\omega \leq \frac{1}{62} \Rightarrow \frac{53}{84} < \frac{19}{16} - 34\omega \Rightarrow \Phi(x+p,\mu) < \Phi(x+p+h,\mu)$.

Thus the center of the ellipsoid has a value of Φ less than any point on the boundary. By applying the strict convexity of Φ we see that the minimizer $x(\mu)$ must lie in the interior of the ellipsoid, proving the lemma. □

The requirement that $2\omega\sqrt{1+\omega} \leq \|p\|_{M/\mu}$ does not significantly complicate the algorithm since ω is fairly small. The continuity of the gradient and Hessian imply that if in the course of the line search this condition fails ($\|p\|_{M/\mu}$ is too small), one can "back out" along the Newton direction to find a point that does satisfy the conditions of the lemma.

Lemma IV.3: If $\omega \leq \frac{1}{62}$ and $\|p\|_{M/\mu} > \frac{1}{8}\sqrt{1-\omega}$, the decrease Δ in Φ after a linesearch along p using Φ as a merit function is

$$\Delta \geq \frac{1}{148}.$$

Proof: Let λ be a step length such that $\|\lambda p\|_H < 1$ and $0 < \lambda \leq 1$. From Lemma IV.1 and the definition of $\hat{q}_x(x + \lambda p, \mu)$,

$$\Phi(x + \lambda p, \mu) \leq \hat{q}_x(x + \lambda p, \mu) + \tfrac{1}{2}\lambda^2\omega\|p\|_H^2 + \frac{\lambda^3\|p\|_H^3}{3(1-\lambda\|p\|_H)}$$

$$= \Phi(x, \mu) + \lambda\nabla\Phi(x)^T p + \tfrac{1}{2}\lambda^2 p^T\tfrac{1}{\mu}M\, p + \tfrac{1}{2}\lambda^2\omega\|p\|_H^2 + \frac{\lambda^3\|p\|_H^3}{3(1-\lambda\|p\|_H)}$$

$$\Rightarrow \quad \Phi(x, \mu) - \Phi(x + \lambda p, \mu) \geq -\lambda\nabla\Phi(x)^T p - \tfrac{1}{2}\lambda^2 p^T\tfrac{1}{\mu}M\, p$$

$$- \tfrac{1}{2}\lambda^2\omega\|p\|_H^2 - \frac{\lambda^3\|p\|_H^3}{3(1-\lambda\|p\|_H)}$$

Now

$$-\nabla\Phi(x)^T p = \nabla\Phi(x)^T(\tfrac{1}{\mu}M)^{-1}\nabla\Phi(x)$$

$$= \nabla\Phi(x)^T(\tfrac{1}{\mu}M)^{-1}(\tfrac{1}{\mu}M)(\tfrac{1}{\mu}M)^{-1}\nabla\Phi(x) = \|p\|_{M/\mu}^2,$$

so

$$\Phi(x, \mu) - \Phi(x + \lambda p, \mu) \geq \lambda\|p\|_{M/\mu}^2 - \tfrac{1}{2}\lambda^2\|p\|_{M/\mu}^2$$

$$- \tfrac{1}{2}\lambda^2\omega\|p\|_H^2 - \frac{\lambda^3\|p\|_H^3}{3(1-\lambda\|p\|_H)} \qquad (4.6)$$

$$\geq (\lambda - \tfrac{1}{2}\lambda^2)(1-\omega)\|\mathrm{p}\|_H^2 - \tfrac{1}{2}\lambda^2\omega\|\mathrm{p}\|_H^2 - \frac{\lambda^3\|\mathrm{p}\|_H^3}{3(1-\lambda\|\mathrm{p}\|_H)} \quad \text{(from (4.2))} \quad (4.7)$$

$$\geq (\lambda - \tfrac{1}{2}\lambda^2)\|\mathrm{p}\|_H^2 - \lambda\omega\|\mathrm{p}\|_H^2 + \tfrac{1}{2}\lambda^2\omega\|\mathrm{p}\|_H^2 - \tfrac{1}{2}\lambda^2\omega\|\mathrm{p}\|_H^2 - \frac{\lambda^3\|\mathrm{p}\|_H^3}{3(1-\lambda\|\mathrm{p}\|_H)}$$

$$\geq (\lambda - \tfrac{1}{2}\lambda^2)\|\mathrm{p}\|_H^2 - \lambda\omega\|\mathrm{p}\|_H^2 - \frac{\lambda^3\|\mathrm{p}\|_H^3}{3(1-\lambda\|\mathrm{p}\|_H)} \quad (4.8)$$

Now let $\lambda = \frac{1}{9\|\mathrm{p}\|_H}$. Then with some manipulation (4.8) becomes

$$\Phi(\mathrm{x}, \mu) - \Phi(\mathrm{x} + \lambda\mathrm{p}, \mu) \geq \tfrac{1-\omega}{9}\|\mathrm{p}\|_H - \tfrac{1}{162} - \tfrac{1}{1944}$$

We also have $\omega \leq \frac{1}{62}$ and $\|\mathrm{p}\|_H > \frac{\sqrt{1-\omega}}{8\sqrt{1+\omega}}$ which together imply

$$\Phi(\mathrm{x}, \mu) - \Phi(\mathrm{x} + \lambda\mathrm{p}, \mu) > \frac{1}{148}.$$

Thus the lemma is proved. □

Corollary IV.3: If $\omega \leq \frac{1}{62}$ and $\|\mathrm{p}\|_{M/\mu} \leq \frac{1}{8}\sqrt{1-\omega}$, then $\Phi(\mathrm{x}, \mu) - \Phi(\mathrm{x} + \mathrm{p}, \mu) \geq 0$

Proof: From (4.8) and $\lambda = 1$ we have

$$\Phi(x, \mu) - \Phi(x + p, \mu) \geq \tfrac{1}{2} \|p\|_H^2 - \omega \|p\|_H^2 - \frac{\|p\|_H^3}{3(1 - \|p\|_H)}$$

$$\geq \left(\tfrac{1}{2} - \omega\right) \|p\|_H^2 - \frac{\frac{1}{8}\|p\|_H^2}{3(\frac{7}{8})}$$

$$\geq \left(\tfrac{1}{2} - \omega - \tfrac{1}{21}\right) \|p\|_H^2 \geq 0. \; \square$$

The next lemma may be of general interest, since it allows one to bound the difference between the current value of $\Phi(x,\mu)$ and the minimum value $\Phi(x(\mu), \mu)$.

Lemma IV.4: If $\omega \leq \frac{1}{62}$ and $2\omega\sqrt{1+\omega} \leq \|p\|_{M/\mu} \leq \frac{1}{8}\sqrt{1-\omega}$, then $\Phi(x, \mu) - \Phi(x(\mu), \mu) \leq 4 \|p\|_H^2$.

Proof: Let $d = x(\mu) - x$. From Lemma IV.1 and the definition of $\hat{q}_x$,

$$\Phi(x(\mu), \mu) \geq \hat{q}_x(x + d, \mu) - \tfrac{1}{2}\omega\|d\|_H^2 - \frac{\|d\|_H^3}{3(1 - \|d\|_H)}$$

$$= \Phi(x, \mu) + \nabla\Phi(x, \mu)^T d + \tfrac{1}{2} d^T \tfrac{1}{\mu} M d - \tfrac{1}{2}\omega\|d\|_H^2 - \frac{\|d\|_H^3}{3(1 - \|d\|_H)}$$

$$= \Phi(x, \mu) - p^T \tfrac{1}{\mu} M d + \tfrac{1}{2} d^T \tfrac{1}{\mu} M d - \tfrac{1}{2}\omega\|d\|_H^2 - \frac{\|d\|_H^3}{3(1 - \|d\|_H)}$$

From the Cauchy-Schwartz inequality we obtain $p^T \frac{1}{\mu} M d \leq \|p\|_{M/\mu} \|d\|_{M/\mu}$. Thus

$$\Phi(x(\mu), \mu) \geq \Phi(x, \mu) - \|p\|_{M/\mu}\|d\|_{M/\mu} + \tfrac{1}{2}\|d\|_{M/\mu}^2 - \tfrac{1}{2}\omega\|d\|_H^2 - \frac{\|d\|_H^3}{3(1 - \|d\|_H)}$$

$$\Phi(x, \mu) - \Phi(x(\mu), \mu) \leq \|p\|_{M/\mu}\|d\|_{M/\mu} - \tfrac{1}{2}\|d\|_{M/\mu}^2 + \tfrac{1}{2}\omega\|d\|_H^2 + \frac{\|d\|_H^3}{3(1 - \|d\|_H)}$$

$$\leq (1+\omega)\|\mathrm{p}\|_H\|\mathrm{d}\|_H - \tfrac{1}{2}(1-\omega)\|\mathrm{d}\|_H^2 + \tfrac{1}{2}\omega\|\mathrm{d}\|_H^2 + \frac{\|\mathrm{d}\|_H^3}{3(1-\|\mathrm{d}\|_H)} \qquad (4.9)$$

where (4.9) follows from (4.2). From the definition of d and Lemma IV.2 we obtain

$$\Phi(\mathrm{x},\,\mu) - \Phi(\mathrm{x}(\mu),\,\mu) \leq \tfrac{5}{2}(1+\omega)\|\mathrm{p}\|_H^2 - \tfrac{1}{2}\|\mathrm{d}\|_H^2 + \omega\|\mathrm{d}\|_H^2 + \frac{\|\mathrm{d}\|_H^3}{3(1-\|\mathrm{d}\|_H)} \qquad (4.10)$$

Again applying Lemma IV.2, the fact that $-\frac{1}{2}+\omega \leq 0$, and $\|\mathrm{p}\|_H \leq \frac{1}{8}$, we have (4.11); (4.12) results from $1+\omega \leq \frac{6}{5}$.

$$\Phi(\mathrm{x},\,\mu) - \Phi(\mathrm{x}(\mu),\,\mu) \leq \tfrac{5}{2}(1+\omega)\|\mathrm{p}\|_H^2 + \frac{\left(\frac{5}{2}\right)^3\|\mathrm{p}\|_H^2}{24\left(\frac{11}{16}\right)}$$

$$\leq \tfrac{5}{2}(1+\omega)\|\mathrm{p}\|_H^2 + \|\mathrm{p}\|_H^2 \qquad (4.11)$$

$$\leq 4\,\|\mathrm{p}\|_H^2 \qquad (4.12)$$

This proves the lemma. □

Corollary IV.3 further shows

$$\Phi(\mathrm{x}+\mathrm{p},\,\mu) - \Phi(\mathrm{x}(\mu),\,\mu) \leq \Phi(\mathrm{x},\,\mu) - \Phi(\mathrm{x}(\mu),\,\mu) \leq 4\,\|\mathrm{p}\|_H^2. \qquad (4.13)$$

Lemma IV.5: If $\omega \leq \frac{1}{62}$ and $2\omega\sqrt{1+\omega} \leq \|p\|_{M/\mu} \leq \frac{1}{8}\sqrt{1-\omega}$, then $\left|b^T x - b^T x(\mu)\right| \leq \frac{1}{2}\mu\sqrt{m}$.

Proof: (See den Hertog, Roos, and Terlaky(1990).) At $x(\mu)$, $\nabla\Phi(x(\mu), \mu) = 0$. Taking the inner product of $\nabla\Phi(x(\mu), \mu)$ with $(x - x(\mu))$ we have

$$\frac{b^T x - b^T x(\mu)}{\mu} - \sum_{i=1}^{m} \frac{\nabla g_i(x(\mu))^T (x - x(\mu))}{g_i(x(\mu))} = 0.$$

Now g_i is concave $\Rightarrow g_i(x) - g_i(x(\mu)) \leq \nabla g_i(x(\mu))^T (x - x(\mu))$

$$\Rightarrow \qquad \frac{b^T x - b^T x(\mu)}{\mu} - \sum_{i=1}^{m} \frac{g_i(x) - g_i(x(\mu))}{g_i(x(\mu))} \geq 0 \tag{4.14}$$

We apply Lemma IV.4 to obtain

$$\Phi(x, \mu) - \Phi(x(\mu), \mu) = \frac{b^T x - b^T x(\mu)}{\mu} - \sum_{i=1}^{m} ln(g_i(x)) - ln(g_i(x(\mu)))$$

$$= \frac{b^T x - b^T x(\mu)}{\mu} - \sum_{i=1}^{m} ln\left(\frac{g_i(x))}{(g_i(x(\mu)}\right)$$

$$\leq 4\|p\|_H^2 \leq \frac{1}{16} \tag{4.15}$$

Define $v_i \equiv g_i(x)/g_i(x(\mu))$ and $w \equiv \frac{b^T(x - x(\mu))}{\mu}$. Then from (4.14) and (4.15)

$$w - \sum_{i=1}^{m}(v_i - 1) = w - \left(\sum_{i=1}^{m} v_i\right) + m \geq 0$$

$$\sum_{i=1}^{m} v_i - \mathrm{m} \leq w \tag{4.16}$$

$$w \leq \frac{1}{16} + \sum_{i=1}^{m} ln(v_i) \tag{4.17}$$

and combining (4.16) and (4.17) we obtain

$$\sum_{i=1}^{m} v_i - \mathrm{m} \leq w \leq \frac{1}{16} + \sum_{i=1}^{m} ln(v_i). \tag{4.18}$$

The geometric-arithmetic mean inequality implies

$$\left(\prod_{i=1}^{m} v_i\right)^{1/m} \leq \frac{1}{m}\sum_{i=1}^{m} v_i .$$

Let $\overline{v} \equiv \frac{1}{m}\sum_{i=1}^{m} v_i$. We can then obtain

$$\mathrm{m}\overline{v} - \mathrm{m} \leq w \leq \frac{1}{16} + ln\prod_{i=1}^{m} v_i = \frac{1}{16} + \mathrm{m}ln\left[\left(\prod_{i=1}^{m} v_i\right)^{1/m}\right] \leq \frac{1}{16} + \mathrm{m}ln(\overline{v})$$

$$\Rightarrow \qquad \overline{v} - 1 \leq \frac{w}{\mathrm{m}} \leq \frac{1}{16\mathrm{m}} + ln(\overline{v}) \tag{4.19}$$

$\frac{d}{d\overline{v}}(ln(\overline{v}) - \overline{v}) = \frac{1}{\overline{v}} - 1 \leq 0$ for $\overline{v} \geq 1$, and $\frac{1}{\overline{v}} - 1 \geq 0$ for $0 < \overline{v} \leq 1$, so $ln(\overline{v}) - \overline{v}$ is monotonically decreasing for $\overline{v} \geq 1$ and monotonically increasing for $0 < \overline{v} \leq 1$.

Suppose $\overline{v} \geq 1 + \frac{1}{2\sqrt{m}}$

$$\Rightarrow \qquad \frac{1}{16\mathrm{m}} + ln(\overline{v}) - \overline{v} + 1 \leq \frac{1}{16\mathrm{m}} + ln\left(1 + \frac{1}{2\sqrt{m}}\right) - \left(1 + \frac{1}{2\sqrt{m}}\right) + 1.$$

The third order Taylor expansion of $ln\left(1+\frac{1}{2\sqrt{m}}\right)$ from 1 is

$$ln\left(1+\frac{1}{2\sqrt{m}}\right)=\frac{1}{2\sqrt{m}}-\frac{1}{8m}+\frac{1}{24\alpha^3 m^{3/2}} \text{ for some } \alpha\in\left[1,\ 1+\frac{1}{2\sqrt{m}}\right].$$

The last term is maximized by $\alpha = 1$, so

$$\frac{1}{16\mathrm{m}}+ln(\overline{v}\,)-\overline{v}\ +1\ \leq\ \frac{1}{16\mathrm{m}}+\frac{1}{2\sqrt{m}}-\frac{1}{8m}+\frac{1}{24m^{3/2}}-\frac{1}{2\sqrt{m}}=$$
$$-\frac{1}{16\mathrm{m}}+\frac{1}{24m^{3/2}}$$

$\mathrm{m}\geq 1 \Rightarrow \frac{1}{16\mathrm{m}}+ln(\overline{v}\,)-\overline{v}\ +1<0$ which contradicts (4.19).

So the supposition is false, and $\overline{v}\ <1+\frac{1}{2\sqrt{m}}$. Now suppose $0<\overline{v}\ \leq\ 1-\frac{1}{2\sqrt{m}}$. We have shown that

$ln(\overline{v}\,)-\overline{v}$ is monotonically increasing in this region. By similarly manipulating the third order Taylor

expansion of $ln\left(1-\frac{1}{2\sqrt{m}}\right)$ we again can show $\frac{1}{16\mathrm{m}}+ln(\overline{v}\,)-\overline{v}\ +1<0$, producing the same

contradiction. Thus $1-\frac{1}{2\sqrt{m}}<\ \overline{v}\ <\ 1+\frac{1}{2\sqrt{m}}$ which leads to the following:

$$\left(1-\frac{1}{2\sqrt{m}}\right)-1\leq\frac{w}{\mathrm{m}}\ \Rightarrow\ -\tfrac{1}{2}\sqrt{\mathrm{m}}\ \leq w$$

$$\frac{w}{\mathrm{m}} \leq \frac{1}{16\mathrm{m}} + ln(\overline{v}) \leq \frac{1}{16\mathrm{m}} + ln\left(1 + \frac{1}{2\sqrt{m}}\right) \leq -\frac{1}{16\mathrm{m}} + \frac{1}{24m^{3/2}} + \frac{1}{2\sqrt{m}} \leq \frac{1}{2\sqrt{m}}$$

$$\Rightarrow \qquad w \leq \tfrac{1}{2}\sqrt{\mathrm{m}} \Rightarrow |w| = \tfrac{1}{\mu}\left(\mathrm{b}^{\mathrm{T}}(\mathrm{x} - \mathrm{x}(\mu))\right) \leq \tfrac{1}{2}\sqrt{\mathrm{m}}$$

$$\Rightarrow \qquad \left|\mathrm{b}^{\mathrm{T}}(\mathrm{x} - \mathrm{x}(\mu))\right| \leq \tfrac{1}{2}\mu\sqrt{\mathrm{m}}\ .\ \text{which proves the lemma. } \square$$

Convergence and Complexity

Theorem IV.1: The algorithm requires at most $\mathrm{K} = \mathrm{O}\left(\frac{1}{\theta} ln\left(\frac{\mathrm{m}}{\epsilon}\right)\right)$ outer iterations to arrive at a solution within an ϵ of the optimal solution value for P.

Proof: $\mu^k = (1-\theta)^k \mu^0$. From duality, we have $\mathrm{b}^{\mathrm{T}}\mathrm{x}(\mu^k) - z^* \leq \mu^k \mathrm{m}$. With Lemma IV.5, this gives us

$$\mathrm{b}^{\mathrm{T}}\mathrm{x}^k - z^* = \mathrm{b}^{\mathrm{T}}\mathrm{x}^k - \mathrm{b}^{\mathrm{T}}\mathrm{x}(\mu^k) + \mathrm{b}^{\mathrm{T}}\mathrm{x}(\mu^k) - z^*$$

$$\leq \left|\mathrm{b}^{\mathrm{T}}\mathrm{x}^k - \mathrm{b}^{\mathrm{T}}\mathrm{x}(\mu^k)\right| + \mathrm{b}^{\mathrm{T}}\mathrm{x}(\mu^k) - z^*$$

$$\leq \tfrac{1}{2}\mu^k\sqrt{\mathrm{m}} + \mu^k\mathrm{m} = \mu^k\left(\tfrac{1}{2}\sqrt{\mathrm{m}} + \mathrm{m}\right)$$

So we need a value of K such that $(1-\theta)^k \mu^0\left(\frac{1}{2}\sqrt{\mathrm{m}} + \mathrm{m}\right) \leq \epsilon$. Taking logarithms, we obtain

$$\mathrm{K} \geq \frac{-ln(\epsilon) + ln\left(\frac{1}{2}\sqrt{\mathrm{m}} + \mathrm{m}\right) + ln(\mu^0)}{-ln(1-\theta)}$$

We have assumed $\mu^0 \leq 1/\epsilon$, and know $\theta \leq -ln(1-\theta)$ for $0 < \theta < 1$. This implies

$$\frac{-2ln(\epsilon) + ln(2\mathrm{m})}{\theta} \geq \frac{-ln(\epsilon) + ln\left(\frac{1}{2}\sqrt{\mathrm{m}} + \mathrm{m}\right) + ln(\mu^0)}{-ln(1-\theta)}$$

Therefore $\mathrm{K} \geq \frac{1}{\theta}(-2ln(\epsilon) + ln(\mathrm{m}) + ln(2))$ satisfies this inequality, proving

$$\mathrm{K} = \mathrm{O}\left(\frac{1}{\theta} ln\left(\frac{\mathrm{m}}{\epsilon}\right)\right). \square$$

Theorem IV.2: The total number of inner iterations, say N, during an arbitrary outer iteration satisfies

$$\mathrm{N}\delta \leq 1 + \frac{\theta}{1-\theta}(\theta\mathrm{m} + \sqrt{\mathrm{m}}), \text{ where } \delta = \frac{1}{148} \text{ is from Lemma IV.3.}$$

Proof: Let $\overline{\mu}$ be the value of the parameter for the current iteration, and μ the value for the previous iteration. Recall $\overline{\mu} = (1-\theta)\mu$. We denote the sequence of points generated by line searches during this iteration as x^0, x^1, x^2, ... x^N. x^0 is the value of x at the beginning of this outer iteration. From Lemma IV.3,

$$\Phi(\mathrm{x}^N, \overline{\mu}) \leq \Phi(\mathrm{x}^0, \overline{\mu}) - \mathrm{N}\delta.$$

Also, $\Phi(\mathrm{x}, \overline{\mu}) = \frac{b^T\mathrm{x}}{(1-\theta)\mu} - \sum_{i=1}^{\mathrm{m}} \ln(g_i(\mathrm{x})) = \frac{b^T\mathrm{x}}{\mu} - \sum_{i=1}^{\mathrm{m}} \ln(g_i(\mathrm{x})) + \frac{\theta b^T\mathrm{x}}{\overline{\mu}}$

$$\Rightarrow \qquad \mathrm{N}\delta \leq \Phi(\mathrm{x}^0, \overline{\mu}) - \Phi(\mathrm{x}^N, \mu) - \frac{\theta b^T\mathrm{x}}{\overline{\mu}}$$

$$\leq \Phi(x^0, \mu) - \Phi(x^N, \mu) + \frac{\theta}{\overline{\mu}}(b^T x^0 - b^T x^N)$$

Because x^0 is almost centered with respect to μ, we have $\Phi(x^0, \mu) - \Phi(x(\mu), \mu) \leq 1$.

$$\Rightarrow \qquad \Phi(x^0, \mu) - \Phi(x^N, \mu) = \Phi(x^0, \mu) - \Phi(x(\mu), \mu) + \Phi(x(\mu), \mu) - \Phi(x^N, \mu)$$

$$\leq 1 + \Phi(x(\mu), \mu) - \Phi(x^N, \mu)$$

$$\leq 1$$

$$\Rightarrow \qquad N\delta \leq 1 + \frac{\theta}{\overline{\mu}}(b^T x^0 - b^T x^N) \tag{4.20}$$

Since x^0 is nearly centered for μ and x^N is nearly centered for $\overline{\mu}$, Lemma IV.5 implies

$$\left|b^T(x^0 - x(\mu))\right| \leq \tfrac{1}{2}\mu\sqrt{m} \tag{4.21}$$

$$\left|b^T(x^N - x(\overline{\mu}))\right| \leq \tfrac{1}{2}\overline{\mu}\sqrt{m}. \tag{4.22}$$

From Lemma III.3, $0 < b^T x(\mu) - b^T x(\overline{\mu})$ and $b^T x(\mu) - m\mu \leq b^T x(\overline{\mu}) - m\overline{\mu}$.

$$\Rightarrow \qquad 0 < b^T x(\mu) - b^T x(\overline{\mu}) \leq m(\mu - \overline{\mu}). \tag{4.23}$$

Combining (4.21), (4.22), and (4.23, we get

$$b^{\mathrm{T}}\mathrm{x}^0 - b^{\mathrm{T}}\mathrm{x}^N = b^{\mathrm{T}}\mathrm{x}^0 - b^{\mathrm{T}}\mathrm{x}(\mu) + b^{\mathrm{T}}\mathrm{x}(\mu) - b^{\mathrm{T}}\mathrm{x}(\overline{\mu}\ ) + b^{\mathrm{T}}\mathrm{x}(\overline{\mu}\ ) - b^{\mathrm{T}}\mathrm{x}^N$$

$$\leq \tfrac{1}{2}\mu\sqrt{\mathrm{m}} + \mathrm{m}(\mu - \overline{\mu}\ ) + \tfrac{1}{2}\overline{\mu}\ \sqrt{\mathrm{m}} \ = \ \mu\Big(\tfrac{1}{2}\sqrt{\mathrm{m}} + \mathrm{m}(1-(1-\theta)) + \tfrac{1}{2}(1-\theta)\sqrt{\mathrm{m}}\Big)$$

$$\leq \mu(\sqrt{\mathrm{m}} + \theta\mathrm{m})$$

Substituting in (4.20) we thus obtain

$$N\delta \ \leq \ 1 + \frac{\theta}{\mu}(b^{\mathrm{T}}\mathrm{x}^0 - b^{\mathrm{T}}\mathrm{x}^N) \leq 1 + \frac{\theta}{1-\theta}(\sqrt{\mathrm{m}} + \theta\mathrm{m}).\ \square$$

Combining Theorems IV.1 and IV.2, the total number of inner iterations is

$$\frac{1}{\delta}\Big(1 + \frac{\theta}{1-\theta}(\sqrt{\mathrm{m}} + \theta\mathrm{m})\Big)\,\mathrm{O}\Big(\frac{1}{\theta}ln\Big(\frac{\mathrm{m}}{\epsilon}\Big)\Big) \ = \ \frac{1}{\delta}\Big(\frac{1}{\theta} + \frac{1}{1-\theta}(\sqrt{\mathrm{m}} + \theta\mathrm{m})\Big)\,\mathrm{O}\Big(ln\Big(\frac{\mathrm{m}}{\epsilon}\Big)\Big).$$

Furthermore, the order of θ has a key role in determining the bound on the number of iterations.

$$\theta = \mathrm{O}\Big(\frac{1}{\sqrt{\mathrm{m}}}\Big) \ \Rightarrow \ \mathrm{O}\Big(\sqrt{\mathrm{m}}\,\big|ln(\epsilon)\big|\Big) \text{ iterations}$$

$$\theta = \mathrm{O}(1) \ \Rightarrow \ \mathrm{O}\Big(\mathrm{m}\,\big|ln(\epsilon)\big|\Big) \text{ iterations}$$

Since each iteration requires $O(n^3)$ computations (the Hessian is inverted), the computational complexity is:

$$O\left(n^3\sqrt{m}\,|ln(\epsilon)|\right) \text{ for } \theta = O\left(\frac{1}{\sqrt{m}}\right)$$

$$O\left(n^3 m\,|ln(\epsilon)|\right) \text{ for } \theta = O(1).$$

Consider the special case where all constraints are linear affine. (This implies P is a linear program.) Then by letting $\epsilon = 2^{-L}$ the purification algorithm for finding an optimal vertex described in Chapter II and in Gonzaga(1992) may be used. Since the complexity of the purification algorithm is the same as that of the Hessian inversion, the overall complexity of the algorithm is not altered.

CHAPTER V: THE PRIMAL-DUAL NEWTON DIRECTION IN THE DUAL VARIABLES

V.1 Analysis of the Dual Variable Newton Direction

In the previous chapter we found that when the deviation from perturbed complementary slackness is kept small, a primal variable algorithm using the primal portion of the Primal-Dual Newton direction will converge to within an ϵ of the solution to problem P in polynomial time. It is therefore logical to consider next the changes in the dual variables induced by a Primal-Dual Newton move. Indeed, if the dual variables change in a manner which maintains the small deviation from perturbed complementary slackness, the polynomality of the Primal Variable Algorithm will also apply to the Primal-Dual Algorithm. In this chapter we investigate the nature of the dual variable movement and how it affects perturbed complementary slackness. It will be shown that the dual variable Newton direction with a standard step size procedure does not guarantee maintenance of a small deviation from perturbed complementary slackness in all dual variables. However, further analysis yields a modification to the step size

which does maintain this small deviation.

Recall the Primal-Dual Newton move:

$$\begin{bmatrix} x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} p \\ -UG^{-1}\nabla g^T p + G^{-1}e\mu - u \end{bmatrix} t$$

where $p \equiv -M^{-1}\left[b^T - \mu\sum_{i=1}^{m} \nabla g_i(x)/g_i(x)\right] = -\left(\frac{1}{\mu}M\right)^{-1}\nabla\Phi(x,\mu)$.

Let $i \in I$ be arbitrary, and $u_i(t)$ denote the new value of the associated dual variable given step size t. Then we have

$$u_i(t) = u_i + \left[-\frac{u_i}{g_i(x)}\nabla g_i(x)^T p + \frac{\mu}{g_i(x)} - u_i\right] t$$

$$\Rightarrow \qquad u_i(t) = u_i\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right) + \left(\frac{\mu}{g_i(x)} - u_i\right) t.$$

The first term of this expression shows how $u_i(t)$ is influenced by tp, the Newton step in the primal

variables. $\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)$ is a first order Taylor series approximation of $g_i(x)/g_i(x+tp)$.

g_i is concave $\Rightarrow 1/g_i$ is convex; and since $g_i(x) > 0 \quad \Rightarrow g_i(x)/g_i(x+tp)$, a function

of t, is convex.

$$\Rightarrow \qquad \frac{g_i(x)}{g_i(x+tp)} \geq 1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp$$

So multiplying u_i by $\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)$ attempts, using first order information, to account for the

change in g_i as x goes to $x + tp$, keeping $u_i g_i(x)$ and $u_i(t) g_i(x + tp)$ approximately equal. However, the inequality above shows the approximation may underestimate $g_i(x)/g_i(x + tp)$, that is,

$$u_i\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right) g_i(x + tp) \leq u_i g_i(x).$$

This underestimation may be observed by assuming a linear affine $g_i(x)$.
$\Rightarrow g_i(x + tp) = g_i(x) + \nabla g_i(x)^T tp$. Substituting into the formula for $u_i(t) g_i(x + tp)$,

$$u_i\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right) g_i(x + tp)$$

$$= u_i g_i(x)\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)\left(1 + \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)$$

$$= u_i g_i(x)\left(1 - \left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)^2\right)$$

$$= u_i g_i(x) - u_i g_i(x)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)^2$$

Since $u_i > 0$ and $g_i(x) > 0$, the last term is always non-positive, and (assuming

$t > 0$) equals zero only if $\nabla g_i(x)^T p = 0$. So for linear affine $g_i(x)$, the underestimation occurrs whenever $\nabla g_i(x)^T p \neq 0$.

More will be said later about how this underestimation could affect the performance of the Primal-Dual Algorithm, but for now let us move on to the second term of

$$u_i(t) = u_i\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T tp\right) + \left(\frac{\mu}{g_i(x)} - u_i\right)t.$$

To study the role of the term $\left(\frac{\mu}{g_i(x)} - u_i\right)t$ in achieving perturbed complementary slackness, we let $p = 0 \Rightarrow g_i(x + tp) = g_i(x)$.

$$\Rightarrow \qquad u_i(t)g_i(x + tp) - \mu = \left(u_i + \left(\frac{\mu}{g_i(x)} - u_i\right)t\right)g_i(x) - \mu$$

$$= \left((1 - t)u_i + \frac{\mu t}{g_i(x)}\right)g_i(x) - \mu$$

$$= (1 - t)u_i g_i(x) - (1 - t)\mu$$

$$= (1 - t)(u_i g_i(x) - \mu)$$

The second term seeks to decrease $|u_i g_i(x) - \mu|$ by a factor of $(1 - t)$; $t = 1$ reduces $|u_i g_i(x) - \mu|$ to zero. But this is predicated on the first term adjusting the value of $u_i(t)$ to account for the change in g_i, keeping $u_i(t)g_i(x + tp) = u_i g_i(x)$. Thus the terms work together: the second term attacks the current deviation from perturbed complementary slackness, and the first adjusts the value of u_i such that

$u_i(t)g_i(x+tp) \doteq u_i g_i(x)$ so as not to confound the effect of the second term.

V.2 The Effect of the Dual Newton Direction on Convergence

To study how the dual variable Newton step performs in reducing $|u_i g_i(x) - \mu|$, we use the value of the step length t introduced in the proof of Lemma IV.3, $t = 1/(9\|p\|_H)$. Besides guaranteeing a reduction in $\Phi(x, \mu)$, this value of λ has other valuable properties:

$$t = 1/(9\|p\|_H) \Rightarrow \|tp\|_H = \tfrac{1}{9} \Rightarrow tp^T H tp = \tfrac{1}{81}. \tag{5.1}$$

The definition of H implies

$$tp^T H tp = tp^T \left[\sum_{i=1}^{m} \frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right] tp = \tfrac{1}{81}. \tag{5.2}$$

All the matrices in this summation in (5.2) are positive semi-definite, so for an arbitrary $i \in I$, we have

$$\left(\frac{1}{g_i(x)} \nabla g_i(x)^T tp \right)^2 \leq \tfrac{1}{81} \tag{5.3}$$

$$\left| \frac{1}{g_i(x)} \nabla g_i(x)^T tp \right| \leq \tfrac{1}{9} \tag{5.4}$$

$$0 \leq -\frac{tp^T \nabla^2 g_i(x)\ tp}{g_i(x)} \leq \frac{1}{81} \tag{5.5}$$

$$0 \leq -\frac{tp^T \nabla^2 g_i(x)\ tp}{2g_i(x)} \leq \frac{1}{162} \tag{5.6}$$

$$0 \leq \left(\frac{1}{g_i(x)} \nabla g_i(x)^T tp\right)^2 - \frac{tp^T \nabla^2 g_i(x)\ tp}{g_i(x)} \leq \frac{1}{81}. \tag{5.7}$$

We are given that $g_i(x)$ is a concave quadratic function. Using the definition of $u_i(t)$ and a 2° Taylor series expansion of $g_i(x + tp)$,

$$u_i(t) g_i(x + tp) - \mu = (u_i g_i(x) - \mu)\left(1 - t\left(1 + \frac{1}{g_i(x)} \nabla g_i(x)^T tp + \frac{tp^T \nabla^2 g_i(x)\ tp}{2g_i(x)}\right)\right)$$

$$+ u_i g_i(x)\left[\frac{tp^T \nabla^2 g_i(x)\ tp}{2g_i(x)} - \left(\frac{1}{g_i(x)} \nabla g_i(x)^T tp\right)^2 - \left(\frac{1}{g_i(x)} \nabla g_i(x)^T tp\right)\left(\frac{tp^T \nabla^2 g_i(x)\ tp}{2g_i(x)}\right)\right]$$

Now define $\omega_i \equiv 1 - \frac{u_i g_i(x)}{\mu} \Rightarrow u_i g_i(x) - \mu = \mu(-\omega_i)$. (Note $\omega_i < 1$ always.) To study the reduction in $\left|u_i g_i(x) - \mu\right|$, consider two cases:

Case 1: $u_i g_i(x) - \mu > 0$. This implies $\omega_i < 0$. Using the inequalities in (5.3) through (5.7) and dropping terms with the appropriate sign, we develope the following inequalities.

$$u_i(t) g_i(x + tp) - \mu \geq (u_i g_i(x) - \mu)\left(1 - t\left(1 + \tfrac{1}{9}\right)\right) + u_i g_i(x)\left(-\tfrac{1}{1458} - \tfrac{1}{81}\right)$$

$$\geq \mu(-\omega_i)\left(1-\frac{10}{9}t\right)+\mu(1-\omega_i)\left(-\frac{19}{1458}\right)$$

Iterations only occur when $\|p\|_H > \frac{1}{8}$, so $t = 1/(9\|p\|_H) \Rightarrow t \leq \frac{8}{9}$. Thus.

$$u_i(t)g_i(x+tp)-\mu \;\geq\; \mu(-\omega_i)\left(1-\frac{80}{81}\right)+\mu\left(-\frac{19}{1458}\right)+\mu(-\omega_i)\left(-\frac{19}{1458}\right)$$

$$\geq\; \mu\omega_i\left(\frac{1}{1458}\right)-\mu\left(\frac{19}{1458}\right)$$

which is clearly a negative lower bound. Now, does this value of t guarantee a reduction in $u_i g_i(x) - \mu$?

$$u_i(t)g_i(x+tp)-\mu \;\leq\; \mu(-\omega_i)\left(1-t\left(1-\frac{1}{9}-\frac{1}{162}\right)\right)+\mu(1-\omega_i)\left(\frac{1}{1458}\right)$$

$$\leq \mu(-\omega_i)\left(1-\frac{143}{162}t+\frac{1}{1458}\right)+\mu\left(\frac{1}{1458}\right)$$

To be certain of a reduction in $u_i g_i(x) - \mu$, we need to show

$$\mu(-\omega_i)\left(1-\frac{143}{162}t+\frac{1}{1458}\right)+\mu\left(\frac{1}{1458}\right) \;<\; \mu(-\omega_i)$$

$$\Rightarrow \qquad -\omega_i\left(-\frac{143}{162}t+\frac{1}{1458}\right) < -\frac{1}{1458}$$

If $0 \leq -\omega_i \leq \frac{1}{62}$, we need $-\frac{143}{162}t+\frac{1}{1458} < -\frac{62}{1458}$, or $t > 0.048951$. So if $\|p\|_H > 2.27$, setting $t = 1/(9\|p\|_H)$ may not give a reduction in $u_i g_i(x) - \mu$.

Case 2: $u_i g_i(x) - \mu \leq 0$. This implies $\omega_i \geq 0$. Using the inequalities in (5.3) through (5.7) and dropping terms with the appropriate sign, we develop the following inequalities.

$$u_i(t)g_i(x+tp) - \mu \leq (u_i g_i(x) - \mu)\left(1 - t\left(1 + \tfrac{1}{9}\right)\right) + u_i g_i(x)\left(\tfrac{1}{1458}\right)$$

$$\leq \mu(-\omega_i)\left(1 - \tfrac{10}{9}t\right) + \mu(1-\omega_i)\left(\tfrac{1}{1458}\right)$$

$$\leq \mu(-\omega_i)\left(1 - \tfrac{10}{9}t + \tfrac{1}{1458}\right) + \mu\left(\tfrac{1}{1458}\right)$$

As before, we know $t \leq \frac{8}{9}$. Thus,

$$u_i(t)g_i(x+tp) - \mu \leq \mu(-\omega_i)\left(1 - \tfrac{80}{81} + \tfrac{1}{1458}\right) + \mu\left(\tfrac{1}{1458}\right)$$

$$\leq \mu(-\omega_i)\left(\tfrac{19}{1458}\right) + \mu\left(\tfrac{1}{1458}\right) \leq \mu\left(\tfrac{1}{1458}\right)$$

which is a quite small positive upper bound for reasonable values of μ. Now, does this value of t guarantee a reduction in $|u_i g_i(x) - \mu|$?

$$u_i(t)g_i(x+tp) - \mu \geq$$
$$\mu(-\omega_i)\left(1 - t\left(1 - \tfrac{1}{9} - \tfrac{1}{162}\right)\right) + \mu(1-\omega_i)\left(-\tfrac{1}{81} - \tfrac{1}{1458}\right)$$

$$\geq \mu(-\omega_i)\left(1 - \tfrac{143}{162}t - \tfrac{19}{1458}\right) - \mu\left(\tfrac{19}{1458}\right)$$

To allow for a reduction in $|u_i g_i(x) - \mu|$, we need to show

$$\mu(-\omega_i)\left(\frac{1439}{1458}-\frac{143}{162}\mathrm{t}\right)-\mu\left(\frac{19}{1458}\right) > \mu(-\omega_i)$$

$$\Rightarrow \qquad \omega_i\left(\frac{143}{162}\mathrm{t}+\frac{19}{1458}\right) > \frac{19}{1458}$$

$$\Rightarrow \qquad \mathrm{t} > \left(\frac{1}{\omega_i}-1\right)(0.014763)$$

A small value of ω_i, say $0 \leq \omega_i \leq \frac{1}{62}$, implies $\mathrm{t} > \frac{8}{9}$. Under the structure of the algorithm, this rules out $\mathrm{t} = 1/(9\|\mathrm{p}\|_H)$, a value of t necessary for the previous convergence and complexity proofs. Therefore in this case, if ω_i satisfies it's upper bound of $\frac{1}{62}$, convergence and polynomial complexity are uncertain. This leads to the content of the next chapter, where analysis yields a modification to the algorithm which allows a certain quantifiable reduction in ω.

CHAPTER VI: MODIFYING THE PRIMAL-DUAL ALGORITHM

VI.1 Modifying the Dual Variable Newton Direction and Step Size

The previous chapter showed the nature of the Primal-Dual Newton direction in the dual variables, and the inherent bias and uncertainty in its effect on *reducing* $|u_i g_i(x) - \mu|$ for each $i \in I$. This chapter develops a modification to the Newton step in the dual variables to remove this bias and uncertainty. We begin with the simpler case of an arbitrary linear affine $g_i(x)$, so that

$$u_i(t) g_i(x + tp) - \mu =$$

$$\left[u_i \left(1 - \frac{1}{g_i(x)} \nabla g_i(x)^T tp \right) + \left(\frac{\mu}{g_i(x)} - u_i \right) t \right] \left(g_i(x) + \nabla g_i(x)^T tp \right) - \mu.$$

A simple way to modify the dual Newton step would be to adjust the step size t (computed in the manner of the Primal Variable Algorithm) using information at hand. Let this adjustment take the form of a parameter τ_i multiplying the step size t in $u_i(t)$. Then the above equation becomes

$$u_i(\tau_i t)g_i(x+tp)-\mu =$$

$$\left[u_i\left(1-\frac{1}{g_i(x)}\nabla g_i(x)^T\tau_i tp\right)+\left(\frac{\mu}{g_i(x)}-u_i\right)\tau_i t\right]\left(g_i(x)+\nabla g_i(x)^T tp\right)-\mu.$$

$$=(u_i g_i(x)-\mu)(1-\tau_i t)-(u_i g_i(x)-\mu)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)\tau_i t$$

$$-u_i g_i(x)\left(\tau_i\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)^2-(1-\tau_i)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)\right)$$

Now let $\tau_i=\bar{\tau}_i\equiv\dfrac{1}{1+\dfrac{1}{g_i(x)}\nabla g_i(x)^T tp}$. With this value we obtain

$$\bar{\tau}_i\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)^2-(1-\bar{\tau}_i)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)=0$$

$$\Rightarrow \quad u_i(\bar{\tau}_i t)g_i(x+tp)-\mu =$$

$$(u_i g_i(x)-\mu)(1-\bar{\tau}_i t)-(u_i g_i(x)-\mu)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)\bar{\tau}_i t$$

$$=(u_i g_i(x)-\mu)\left(1-\bar{\tau}_i t\left(1+\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)\right)$$

$$=(u_i g_i(x)-\mu)(1-t).$$

Thus, multiplying t by $\bar{\tau}_i$ in the computation of the new value of u_i decreases $\left|u_i g_i(x)-\mu\right|$ by a factor of $(1-t)$ at each inner iteration. (Recall $0<t\leq 1$.) This has a powerful implication. Suppose the algorithm is at a point near the trajectory,

that is, $\omega\sqrt{1+\omega} \leq \|p\|_{M/\mu} < \frac{1}{8}\sqrt{1-\omega}$; and $u_i g_i(x) - \mu = 0$, $i \in I$. Let $\overline{\mu} = (1-\theta)\mu$. Then $u_i g_i(x) - \overline{\mu} = u_i g_i(x) - \mu + \theta\mu = \theta\mu$, $i \in I$. If each g_i is linear affine and the algorithm uses the formula $u_i(\overline{\tau}_i t)$, after any given iteration the value of $u_i g_i(x) - \mu$ is the same for all $i \in I$, and $\frac{1}{\mu}M = H - \omega H$, where $\omega = 1 - u_i g_i(x)/\mu$, $i \in I$. So within an arbitrary outer iteration, the initial common value of $u_i g_i(x) - \mu$ is reduced strictly monotonically at each inner iteration. Furthermore, if any inner iteration returns a step size $t = 1$, then $u_i g_i(x) - \mu = 0$, $i \in I$ at each subsequent inner iteration.

This result applies only to linear affine constraints, so the next logical step would be to attempt a similar result for the more general concave quadratic constraint. We begin with the general step size multiplier τ_i used successfully before.

$$u_i(\tau_i t)g_i(x+tp) - \mu = \left[u_i\left(1 - \frac{1}{g_i(x)}\nabla g_i(x)^T \tau_i tp\right) + \left(\frac{\mu}{g_i(x)} - u_i\right)\tau_i t\right]$$

$$\times \left(g_i(x) + \nabla g_i(x)^T tp + \tfrac{1}{2}tp^T \nabla^2 g_i(x) tp\right) - \mu$$

$$= (u_i g_i(x) - \mu)(1 - \tau_i t) - (u_i g_i(x) - \mu)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp + \tfrac{1}{2}tp^T \frac{\nabla^2 g_i(x)}{g_i(x)} tp\right)\tau_i t \qquad (6.1)$$

$$- u_i g_i(x)\left(\tau_i\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)^2 - \tfrac{1}{2}tp^T \frac{\nabla^2 g_i(x)}{g_i(x)} tp - (1-\tau_i)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)\right) \qquad (6.2)$$

$$- u_i g_i(x)\left(\frac{\tau_i}{2}\left(tp^T \frac{\nabla^2 g_i(x)}{g_i(x)} tp\right)\left(\frac{1}{g_i(x)}\nabla g_i(x)^T tp\right)\right) \qquad (6.3)$$

The presence of $\frac{1}{2}\mathrm{tp}^{\mathrm{T}}\frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\mathrm{tp}$ in (6.2) not multiplied by τ_i, coupled with the

possibility that $\nabla g_i(\mathrm{x})^{\mathrm{T}}\mathrm{p} = 0$ means that in general it may not be possible to zero out (6.2) and (6.3) with the correct choice of τ_i. But the components containing τ_i in (6.2) and (6.3) can be made to sum to zero, as follows:

Let $\bar{\tau}_i = \dfrac{1}{1 + \frac{1}{g_i(\mathrm{x})}\nabla g_i(\mathrm{x})^{\mathrm{T}}\mathrm{tp} + \frac{1}{2}\mathrm{tp}^{\mathrm{T}}\frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\mathrm{tp}}$ and the following equations obtain.

$$\bar{\tau}_i\left(\frac{1}{g_i(\mathrm{x})}\nabla g_i(\mathrm{x})^{\mathrm{T}}\mathrm{tp}\right)^2 - (1-\bar{\tau}_i)\left(\frac{1}{g_i(\mathrm{x})}\nabla g_i(\mathrm{x})^{\mathrm{T}}\mathrm{tp}\right)$$

$$+\frac{\bar{\tau}_i}{2}\left(\mathrm{tp}^{\mathrm{T}}\frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\mathrm{tp}\right)\left(\frac{1}{g_i(\mathrm{x})}\nabla g_i(\mathrm{x})^{\mathrm{T}}\mathrm{tp}\right) = 0.$$

$\mathrm{u}_i(\bar{\tau}_i\,\mathrm{t})g_i(\mathrm{x}+\mathrm{tp}) - \mu =$

$$(\mathrm{u}_i g_i(\mathrm{x}) - \mu)\left(1 - \bar{\tau}_i\,\mathrm{t}\left(1 + \frac{1}{g_i(\mathrm{x})}\nabla g_i(\mathrm{x})^{\mathrm{T}}\mathrm{tp} + \frac{1}{2}\mathrm{tp}^{\mathrm{T}}\frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\mathrm{tp}\right)\right)$$

$$-\mathrm{u}_i g_i(\mathrm{x})\left(-\frac{1}{2}\mathrm{tp}^{\mathrm{T}}\frac{\nabla^2 g_i(\mathrm{x})}{g_i(\mathrm{x})}\mathrm{tp}\right)$$

$$= (\mathrm{u}_i g_i(\mathrm{x}) - \mu)(1-\mathrm{t}) + \mathrm{u}_i\left(\frac{1}{2}\mathrm{tp}^{\mathrm{T}}\nabla^2 g_i(\mathrm{x})\mathrm{tp}\right).$$

The use of $\bar{\tau}_i$ has removed some of the uncertainty concerning the reduction of

$\left|u_i g_i(x) - \mu\right|$, but a negative bias remains: $u_i\left(\frac{1}{2}tp^T\nabla^2 g_i(x)tp\right) \leq 0$. However, if we define

$$\bar{\beta}_i \equiv -\frac{u_i\left(\frac{1}{2}tp^T\nabla^2 g_i(x)tp\right)}{1 + \frac{1}{g_i(x)}\nabla g_i(x)^T tp + \frac{1}{2}tp^T\frac{\nabla^2 g_i(x)}{g_i(x)}tp} = -\frac{1}{2}\bar{\tau}_i u_i\left(tp^T\frac{\nabla^2 g_i(x)}{g_i(x)}tp\right)$$

we obtain

$$\left(u_i(\bar{\tau}_i t) + \bar{\beta}_i\right)g_i(x + tp) - \mu = u_i(\bar{\tau}_i t)g_i(x + tp) - \mu + \bar{\beta}_i g_i(x + tp)$$

$$= (u_i g_i(x) - \mu)(1 - t) + u_i\left(\frac{1}{2}tp^T\nabla^2 g_i(x)tp\right) - u_i\left(\frac{1}{2}tp^T\nabla^2 g_i(x)tp\right)$$

$$= (u_i g_i(x) - \mu)(1 - t).$$

The use of $\bar{\tau}_i$ and $\bar{\beta}_i$ in this manner removes uncertainty and bias concerning the reduction of $\left|u_i g_i(x) - \mu\right|$ in the same way as was achieved for the linear affine $g_i(x)$; and when $\nabla^2 g_i(x) = 0$ for all x (g_i is linear affine), $\bar{\beta}_i = 0$ and this technique reduces to the method developed earlier for linear affine $g_i(x)$. The following definitions will allow incorporation of this modification into the Primal-Dual Algorithm:

$$u_i(t, \bar{\tau}_i, \bar{\beta}_i) \equiv u_i + \left[\frac{u_i}{g_i(x)}\nabla g_i(x)^T p + \frac{\mu}{g_i(x)} - u_i\right]\bar{\tau}_i t + \bar{\beta}_i$$

or in vector-matrix notation,

$$u(t, \bar{\tau}, \bar{\beta}) \equiv u + \bar{\tau}\left[-UG^{-1}\nabla g^{T}p + G^{-1}e\mu - u\right]t + \bar{\beta}$$

where $\bar{\beta}$ is an $m \times 1$ vector, the i^{th} entry of which is $\bar{\beta}_i$, and $\bar{\tau}$ is an $m \times m$ matrix equal to $\mathrm{diag}(\bar{\tau}_i)$.

VI.2 A Modified Primal-Dual Algorithm

The Modified Primal-Dual Algorithm (MPDA) is essentially the Primal Variable Algorithm with the addition of the modified Primal-Dual Newton step in the dual variables. Other changes have been made so that the requirement of $\omega \leq 1/62$ is met. Consider the algorithm at the end of an outer iteration for an arbitrary $\mu > 0$, meaning the tolerance $\|p\|_{M/\mu} \leq \tau = \frac{1}{8}\sqrt{1-\omega}$ has been met. Suppose the algorithm has also achieved $u_i g_i(x) - \mu = 0$, $i \in I$. This implies $\omega = 0$, $M/\mu = H$, and the tolerance thus becomes $\|p\|_{M/\mu} = \|p\|_H \leq = \frac{1}{8}$. It is at this point in the algorithm that μ is decremented. Let $\bar{\mu}$ denote the decremented value, so $\bar{\mu} = (1-\theta)\mu$ for some $\theta \in (0, 1)$. When μ is so decremented, ω no longer equals zero. In fact, for an arbitrary $i \in I$,

$$\left|1 - \frac{u_i g_i(x)}{\bar{\mu}}\right| = \left|\frac{1}{1-\theta}\left(1 - \theta - \frac{u_i g_i(x)}{\mu}\right)\right| = \left|\frac{-\theta}{1-\theta}\right| = \frac{\theta}{1-\theta} \Rightarrow \omega = \frac{\theta}{1-\theta}.$$

By specifying $\theta \leq 1/63$, we ensure that when reducing μ to $\bar{\mu} = (1-\theta)\mu$ we have $\omega \leq 1/62$ at the start of the next iteration. Recall that with the modification to the dual Newton step, ω is reduced monotonically during the course of an outer

iteration, and once a step size t=1 is achieved the algorithm maintains $u_i g_i(x) - \mu = 0$, $i \in I$ (and therefore $\omega = 0$) for the remainder of that outer iteration until the tolerance is met and μ is decremented. If the situation occurs where the tolerance is met at the end of an outer iteration, but $u_i g_i(x) - \mu \neq 0$ for some $i \in I$, the algorithm simply takes one more Primal-Dual Newton step with $t = 1$, achieving $u_i g_i(x) - \mu = 0$, $i \in I$, and $\omega = 0$. By Corollary IV.3 this maintains the necessary proximity to the central path. μ is then decremented to $\overline{\mu} = (1 - \theta)\mu$ resulting in $\omega = 1/62$, and the next outer iteration begins.

Input

Reduction factor $\theta \in (0, 1/63]$

Tolerance $\tau = \frac{1}{8}\sqrt{1 - \omega}$. Assume $\max_{i \in I} \left| 1 - \frac{u_i g_i(x)}{\mu} \right| \leq \frac{1}{62}$ always.

Υ is an accuracy parameter.

$x^0 \in \mathbb{R}^*$ is given, as is $\mu^0 \leq 1/\Upsilon$, and $u_i^0 > 0$, $i \in I$ such that $\left\| p(x^0, u^0, \mu^0) \right\|_{M/\mu} \leq \tau$ and $u_i^0 g_i(x^0) - \mu^0 = 0$, $i \in I$

begin

$x = x^0$, $u = u^0$, $\mu = \mu^0$

while $\mu > \Upsilon/2m$ do

begin outer iteration

while $\| p \|_{M/\mu} \leq \tau$ do

begin inner iteration

$$\overline{\lambda} = \arg \min_{0 \leq \lambda \leq 1} \{\Phi(x, \mu) \text{ s.t. } (x + \lambda p) \in \mathbb{R}^*, \; \| p \|_{M/\mu} \geq 2\omega\sqrt{1 + \omega}\}$$

$x = x + \bar{\lambda}\, p$

$u = u(\bar{\lambda}, \bar{\tau}, \bar{\beta})$

end inner iteration

if $\max_{i \in I} \left| 1 - \frac{u_i g_i(x)}{\mu} \right| > 0$

begin one extra step

$x = x + p$

$u = u(1, \bar{\tau}, \bar{\beta})$

end one extra step

$\mu = (1 - \theta)\mu$

end outer iteration

end

The lemmas and convergence/complexity theorems apply to the Modified Primal-Dual Algorithm as they did to the Primal Variable Algorithm. The bounds on the number of iterations two choices of θ are:

$$\theta = O\left(\frac{1}{\sqrt{m}}\right) \Rightarrow O\left(\sqrt{m}\,|ln(\epsilon)|\right) \text{ iterations}$$

$$\theta = O(1) \Rightarrow O\left(m\,|ln(\epsilon)|\right) \text{ iterations}$$

and the corresponding computational complexity is:

$$O\left(n^3\sqrt{m}\,|ln(\epsilon)|\right) \text{ for } \theta = O\left(\frac{1}{\sqrt{m}}\right)$$

$$O\left(n^3 m\,|ln(\epsilon)|\right) \text{ for } \theta = O(1)$$

As stated at the end of Chapter 3, if the problem P has only linear affine constraints, then setting $\epsilon = 2^{-L}$ allows a purification algorithm to find an optimal vertex with no increase in complexity (see Gonzaga(1992)).

Another significant attribute of the Modified Primal-Dual Algorithm is convergence of the Lagrangian multipliers to an optimal vector of multipliers, u^*. This results from the way in which the MPDA assumes the behavior of the standard primal variable SUMT at the end of each outer iteration. Suppose the algorithm is at the end of an outer iteration for an arbitrary $\mu > 0$. Then $u_i g_i(x) = \mu$, $i \in I$. Additional modified primal-dual Newton steps will coincide with standard SUMT Newton steps, and will also maintain $u_i g_i(x) = \mu$, $i \in I$. Thus, the MPDA inherits the Lagrangian multiplier convergence property of SUMT, as proved in Theorems 25 and 26 of Fiacco and McCormick(1968, 1990).

CHAPTER VII: LINEAR PROGRAMMING WITH THE MODIFIED PRIMAL-DUAL ALGORITHM

VII.1 A Review of Anstreicher's Results

In previous chapters we have developed a modification to the Primal-Dual Newton step which allowed for a guaranteed reduction in the deviation from perturbed complementary slackness with each Newton step. That reduction in turn allowed the application of the den Hertog, Roos, and Terlaky(1990) analytical techniques to prove convergence and polynomial complexity for the Modified Primal-Dual Algorithm. This chapter will show that the modificationt also allows us to adapt Anstreicher's results for quadratic programming with the logarithmic barrier function to linear programming with the Modified Primal-Dual Algorithm. Thus, we have another way to show polynomial complexity for the Modified Primal-Dual Algorithm, albeit for a less general class of problems. (An attempt was made to adapt Anstreicher's results directly to the Primal Variable Algorithm for small ω, as was done with the work of den Hertog, Roos, and Terlaky. Because of the relatively tight nature of the inequalities in several of the lemmas, however,

this has not been successful.)

We begin by reviewing Anstreicher's convergence and complexity proofs for the traditional SUMT logarithmic barrier algorithm applied to a quadratic program. Such a review allows the reader to see a significantly different approach to proving polynomality, one that makes important use of the linear algebra associated with the linear affine constraints of the quadratic program. We then show that the Modified Primal-Dual Algorithm, when applied to a linear program, can be analyzed for convergence and complexity in the same manner as Anstreicher's treatment of SUMT. The lemmas and theorems from Anstreicher(1990) are presented in this chapter without proof. For the interested reader, the proofs are in Appendix 1, precisely as developed in Anstreicher(1990) and Anstreicher et al(1990) with only notational changes to conform to this dissertation.

Let us define the standard quadratic program (QP) in E^n using notation similar to Anstreicher:

$$\text{QP: } \min f(x) = b^T x + \tfrac{1}{2} x^T Q x$$
$$\text{s.t. } Ax - c \geq 0$$

for $b \in E^n$, $c \in E^n$, A an $m \times n$ matrix, and Q an $n \times n$ symmetric positive semi-definite matrix. We assume the data is integer with bit length L, and as before the feasible region R is bounded with a non-empty interior $\mathbb{R}^*$. This implies that $m > n$ and the columns of A are linearly independent. The Wolfe dual of QP is

$$\text{QD: } \max b^T x + \tfrac{1}{2} x^T Q x - u^T (Ax - c)$$

$$\text{s.t. } b + Qx - A^T u = 0$$
$$u \geq 0$$

The dual can be simplified. $b + Qx - A^T u = 0 \Rightarrow u^T A = b^T + x^T Q \Rightarrow$

$$b^T x + \tfrac{1}{2} x^T Q x - u^T A x + u^T c = b^T x + \tfrac{1}{2} x^T Q x - (b^T + x^T Q)x + c^T u$$

$$= -\tfrac{1}{2} x^T Q x + c^T u$$

Thus we can write the dual problem as

$$\text{QD: } \max d(x,u) = -\tfrac{1}{2} x^T Q x + c^T u$$
$$\text{s.t. } b + Qx - A^T u = 0$$
$$u \geq 0$$

Lemma VII.1: Let x be feasible in QP and $(\hat{x}, \hat{u})$ be feasible in QD. Then $f(x) - d(\hat{x}, \hat{u}) \geq 0$.

Let a_i^T denote the i^{th} row of A. We define the logarithmic barrier function $\Phi(x, \mu)$ for $x \in \mathbb{R}^*$ and $\mu > 0$ as before.

$$\Phi(x,\mu) = \frac{f(x)}{\mu} - \sum_{i=1}^{m} ln(a_i^T x - c_i)$$

The gradient and Hessian of $\Phi(x,\mu)$ are

$$\nabla\Phi(x,\mu) = \frac{Qx+b}{\mu} - \sum_{i=1}^{m} \frac{a_i}{(a_i{}^T x - c_i)}$$

The n × n matrix H will again denote the Hessian of Phi. By Lemma III.2. H is positive definite. $\Phi(x,\mu)$ is continuous on $\mathbb{R}^*$ and goes to $+\infty$ at the boundary of R, so the compactness of R guarantees $\Phi(x,\mu)$ has a minimizer $x(\mu) \in \mathbb{R}^*$.

For an arbitrary $\mu > 0$, Anstreicher employs three different measures of the distance of a point $x \in \mathbb{R}^*$ from the central path $x(\mu)$. Some notational definitions are necessary: $s \equiv Ax - c$. $S \equiv \text{diag}(s)$, $s^{-1} \equiv (1/s_1, 1/s_2, \ldots, 1/s_m)^T = S^{-1}e$. The primal variable Newton step is $p_x \equiv -H^{-1}\nabla\Phi(x,\mu)$, and $p_s \equiv Ap_x$ is the step in s resulting from the Newton step p_x in x. The three measures are:

(1) $\delta(x,\mu) = \min_{u|\ A^T u = b + Qx} \left\| \frac{1}{\mu} Su - e \right\|$

(2) $\left\| S^{-1} p_s \right\|$

(3) $\| p_x \|_H$

To better understand these measures, it may be helpful to apply them to the familiar problem P from previous chapters: $\left\| \frac{1}{\mu} Su - e \right\|$ is the same as $\left\| u^T g - \mu \right\|$, (3) is the norm of the Newton step in the primal variables with respect to the Hessian of $\Phi(x,\mu)$, and (2) is the same as (3) but ignoring the contribution of the objective function f(x).

Lemma VII.2: Let $x \in \mathbb{R}^*$ and $\mu > 0$ be arbitrary. Then
$\|S^{-1}p_s\|^2 \leq \|p_x\|_H^2 \leq \delta(x,\mu)^2$.

Lemma VII.3: Given $x \in \mathbb{R}^*$, suppose $\|S^{-1}p_s\| < 1$. Let $\bar{x} = x + p_x$, $\bar{s} = s + p_s$, and $\bar{S} = \mathrm{diag}(\bar{s})$. Let $\bar{H} = \nabla^2\Phi(\bar{x},\mu)$, $p_{\bar{x}} = -\bar{H}^{-1}\nabla\Phi(\bar{x},\mu)$, and $p_{\bar{s}} = Ap_{\bar{x}}$. Then $\bar{x} \in \mathbb{R}^*$ and

$$\|\bar{S}^{-1}p_{\bar{s}}\| \leq \|p_{\bar{x}}\|_{\bar{H}}^2 \leq \delta(\bar{x},\mu) \leq \|S^{-1}p_s\|^2 \leq \|p_x\|_H^2 \leq \delta(x,\mu)^2$$

Lemma VII.4: If $x \in \mathbb{R}^*$ and $\|p_x\|_H < 1$, then $\Phi(x,\mu) - \Phi(x(\mu),\mu) \leq \dfrac{\|p_x\|_H^2}{1 - \|p_x\|_H^2}$.

Lemma VII.5: If $x \in \mathbb{R}^*$ and $\|p_x\|_H < 1$, then

$$|f(x) - f(x(\mu))| \leq \frac{\|p_x\|_H(1 + \|p_x\|_H)}{1 - \|p_x\|_H}\mu\sqrt{m}$$

The algorithm used by Anstreicher and Anstreicher et al is similar to that of den Hertog, Roos, and Terlaky(1990). It begins with a feasible interior point x^0 and an initial barrier parameter value $\mu^0 = 2^{O(L)}$. From point x^k, the algorithm conducts linesearches along a sequence of Newton steps until it reaches a point where $\|p_x\|_H \leq \tau$. This point becomes x^{k+1}, μ^k is decremented to $\mu^{k+1} = (1-\theta)\mu^k$ for some fixed $\theta \in (0,1)$, and the process starts over again. The algorithm terminates when $\|p_x\|_H \leq \tau$ and $m\mu^k \leq \epsilon = 2^{-O(L)}$, from which the

optimal solution may be found in $O(m^3)$ operations. Convergence and complexity of the algorithm are proven by Anstreicher with the following two theorems.

Theorem VII.1: Let z^* denote the minimum value of f(x) for problem QP. After $O(L/\theta)$ outer iterations, the algorithm obtains $\mu = \mu^k$ such that $m\mu^k \leq 2^{-O(L)}$ and a point $x = x^{k+1}$ such that $f(x) - z^* \leq \left(1 + \frac{\tau(1+\tau)}{(1-\tau)\sqrt{m}}\right) 2^{-O(L)}$.

The proof of Theorem VII.2 requires the following lemma proving a guaranteed reduction in $\Phi(\cdot,\mu)$ along the Newton direction. It is similar to the den Hertog, Roos, and Terlaky Taylor series expansion result used in the proof of Lemma IV.1. In that case, the proof was quite long and detailed, and was omitted. Anstreicher's proof for this lemma is somewhat shorter, and is included in Appendix 1.

Lemma VII.6: Let $\overline{\lambda} = \frac{1}{1+\|p_x\|_H}$. Then

$$\Phi(x,\mu) - \Phi(x + \overline{\lambda}\, p_x,\mu) \geq \|p_x\|_H - ln(1+\|p_x\|_H)$$

Corollary VII.6: Let $\tau < \|p_x\|_H \leq \frac{1}{2}$. Then a pure Newton step ($\lambda = 1$) obtains

$$\Phi(x,\mu) - \Phi(x + p_x,\mu) \geq \Delta = \tau^2 + ln(1-\tau) + \tau > 0$$

We can now proceed to Theorem VII.2, which bounds the number of inner

iterations per outer iteration.

Theorem VII.2: Let $\Delta = \tau - ln(1+\tau)$, $E_\Phi = \frac{\tau^2}{1-\tau^2}$, $E_f = \frac{\tau(1+\tau)}{1-\tau}$, and N denote the number of inner iterations required for an arbitrary outer iteratation. Then N is bounded as:

$$N \leq \frac{1}{\Delta}\left(\frac{\theta}{1-\theta}\left(\theta m + 2E_f\sqrt{m}\right) + E_\Phi\right)$$

Combining Theorem VII.1 and VII.2 gives the complexity result we seek: the total number of Newton iterations required by the algorithm to solve problem QP is bounded by

$$\frac{1}{\Delta}\left(\frac{\theta m + 2E_f\sqrt{m}}{1-\theta} + \frac{E_\Phi}{\theta}\right) O(L).$$

The choice of θ has much the same influence as in the results of den Hertog, Roos, and Terlaky:

$$\theta = O\left(\frac{1}{\sqrt{m}}\right) \Rightarrow O(\sqrt{m}\ L) \text{ iterations}$$

$$\theta = O(1) \Rightarrow O(mL) \text{ iterations.}$$

In Corollary VII.6 it was shown that a pure Newton step would achieve a guaranteed reduction in $\Phi(\cdot, \mu)$ if $\tau < \|p_x\|_H \leq \frac{1}{2}$. To ensure that $\|p_x\|_H \leq \frac{1}{2}$

throughout the algorithm, note that Theorem VII.2 implies that

$$\frac{\theta}{1-\theta}\left(\theta m + 2E_f\sqrt{m}\right) + E_\Phi$$

is the maximum reduction possible for any step size, while Lemma VII.6 shows a reduction of $0.5 - ln(1.5) > .0945$ is guaranteed with the correct choice of step size for $\|p_x\|_H > \frac{1}{2}$. Thus it suffices to restrict the maximum possible reduction to less than .0945 to ensure that $\|p_x\|_H \leq \frac{1}{2}$. This may be done by choosing θ such that

$$\frac{\theta}{1-\theta}\left(\theta m + 2E_f\sqrt{m}\right) + E_\Phi \leq .0945. \tag{7.1}$$

For small values of τ both E_f and E_Φ approach zero, so θ such that $\theta^2 m \leq .0945$, or $\theta \leq .3/\sqrt{m}$ will suffice. This result will be important in later analysis.

This concludes the review of Anstreicher's complexity results for QP; the remaining portion of this chapter will show that these results may be applied to linear programming with the Modified Primal-Dual Algorithm to achieve the same complexity bound.

VII.2 Another Polynomial Complexity Result for the Modified Primal Dual Algorithm

To remain consistant with the notation used in the development of the Modified Primal-Dual Algorithm, let us redefine the linear program (LP) in a manner identical to the definition of problem P in Chapter III, with the additional

restriction that the constraints be linear affine:

$$\begin{aligned} \text{LP: } & \min\ b^{\mathrm{T}}\mathrm{x} \\ & \text{s.t. } g_i(\mathrm{x}) \geq 0 \text{ and linear affine, } \mathrm{i} \in \mathrm{I} \\ & \mathrm{x} \in \mathrm{E}^n \end{aligned}$$

All restrictions on the feasible region for P apply to LP. This LP is the problem we will address throughout the rest of this chapter.

In Chapter III this thesis introduced the matrix M and its relationship to the Hessian of the logarithmic barrier function. Of particular interest is

$$\tfrac{1}{\mu}\mathrm{M}(\mathrm{x},\mu) = \nabla_x^2\Phi(\mathrm{x},\mu) - \sum_{i=1}^{m}\left(1 - \frac{\mathrm{u}_i g_i(\mathrm{x})}{\mu}\right)\left[\frac{\nabla g_i(\mathrm{x})\nabla g_i(\mathrm{x})^{\mathrm{T}}}{g_i(\mathrm{x})^2}\right],$$

where $\mathrm{u}_i > 0$, $\mathrm{i} \in \mathrm{I}$. As before, let H denote the Hessian of $\Phi(\,\cdot\,,\mu)$. It is clear that if $\mathrm{u}_i g_i(\mathrm{x}) - \mu = 0$ for all $\mathrm{i} \in \mathrm{I}$, then $\mathrm{H}(\mathrm{x},\mu) = \frac{1}{\mu}\mathrm{M}(\mathrm{x},\mu)$.

Lemma VII.7: Let $\mathrm{u}_i g_i(\mathrm{x}) - \mu = 0$ for all $\mathrm{i} \in \mathrm{I}$, and let $\theta \in (0,1)$ be given. If $\overline{\mu} = (1-\theta)\mu$, then

$$\tfrac{1}{\overline{\mu}}\mathrm{M}(\mathrm{x},\overline{\mu}) = \left(\frac{1}{1-\theta}\right)\mathrm{H}(\mathrm{x},\mu)$$

Proof: From (3.7) we have

$$\tfrac{1}{\mu}\mathrm{M}(\mathrm{x},\mu) = \sum_{i=1}^{m}\frac{\mathrm{u}_i g_i(\mathrm{x})}{\mu}\left[\frac{\nabla g_i(\mathrm{x})\nabla g_i(\mathrm{x})^{\mathrm{T}}}{g_i(\mathrm{x})^2}\right].$$

Using $u_i g_i(x) = \mu$ and $\overline{\mu} = (1-\theta)\mu$, we get

$$\frac{1}{\overline{\mu}} M(x,\overline{\mu}) = \sum_{i=1}^{m} \frac{\mu}{(1-\theta)\mu} \left[\frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} \right] = \left(\frac{1}{1-\theta} \right) H(x,\mu)$$

which completes the proof. □

A Linear Programming Version of the MPDA. The algorithm begins with a feasible interior point x^0, an initial barrier parameter value $\mu^0 = 2^{O(L)}$, $u_i^0 > 0$ such that $u_i^0 g_i(x^0) - \mu^0 = 0$ for all $i \in I$, and $\| p \|_{M/\mu} = \| p \|_H \leq \tau$. Let the k^{th} outer iteration be arbitrary. Starting from point x^k, the algorithm takes a series of pure Primal-Dual Newton steps in the primal variables, that is,

$$p_{x_j} = -\left(\frac{M(x_j^k, \mu^k)}{\mu^k} \right)^{-1} \nabla \Phi(x_j^k, \mu^k), \; j = 1, 2, 3, \ldots$$

At each such step, each u_i, $i \in I$ is updated to $u_i(1, \overline{r}_i, 0)$. These inner iterations continue until the algorithm reaches a point where $\| p_x \|_H \leq \tau$, say the $\widehat{j}^{th}$ iteration. Then $x^{k+1} = x_{\widehat{j}}^k$, μ^k is decremented to $\mu^{k+1} = (1-\theta)\mu^k$ for some fixed $\theta \in (0,1)$ satisfying (7.1), and the process starts over again. The algorithm terminates when $\| p_x \|_H \leq \tau$ and $m\mu^k \leq \epsilon = 2^{-O(L)}$.

The initial Primal-Dual Newton step in the primal variables will be identical in direction to that generated by the logarithmic barrier function, but will be shorter by a factor of $(1-\theta)$. Therefore, the reduction Δ in $\Phi(\cdot, \mu)$ from this first iteration

will be

$$\Delta \geq (1-\theta)\tau^2 + ln(1-(1-\theta)\tau) + (1-\theta)\tau. \tag{7.2}$$

Since τ is small and positive, the right hand side of (7.2) is strictly positive for $\theta = 0$, equals zero for $\theta = 1$, and is strictly monotonically decreasing with respect to θ on $(0,1)$. Thus, $\Delta > 0$ for $\theta \in (0,1)$. After the initial pure Primal-Dual Newton step in the primal variables, $M/\mu = H$ and the exact results of Anstreicher are obtained. (It is also permissible to conduct linesearches along the Primal-Dual Newton directions after the first inner iteration's pure Primal-Dual Newton step. The decrease in $\Phi(\cdot, \mu)$ will then be guaranteed by Lemma VII.6 instead of Corollary VII.6.) Therefore, the complexity results proved by Theorems VII.1 and VII.2 for the logarithmic barrier function applied to QP apply for this version of the Modified Primal-Dual Algorithm applied to LP.

CHAPTER VIII: SATISFYING THE INITIAL CONDITION REQUIREMENTS

We begin this chapter by stating the general problem to be addressed and key notation used.

$$\min_{x \in E^n} f(x)$$

$$\text{s.t. } g_i(x) \geq 0,\ i \in I \equiv (1, 2, \ldots, m)$$

where $-f(x)$ and $g_i(x)$, $i \in (1, 2, \ldots, m)$ are concave quadratic functions.

Standard Notation and Definitions

$$\mathbb{R} \equiv \{x \in E^n \mid g_i(x) \geq 0,\ i \in I\}$$

$$\mathbb{R}^* \equiv \{x \in E^n \mid g_i(x) > 0,\ i \in I\}$$

$$\Phi(x,\mu) \equiv \frac{f(x)}{\mu} - \sum_{i=1}^{m} \ln(g_i(x)).$$

$$\nabla\Phi(x,\mu) = \frac{\nabla f(x)}{\mu} - \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)}$$

$$H(x,\mu) \equiv \nabla^2\Phi(x,\mu) = \frac{\nabla^2 f(x)}{\mu} + \sum_{i=1}^{m} \frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}$$

(Note: when f(x) is linear, $\nabla^2 f(x)$ vanishes.)

$$M(x,u,\mu) \equiv \sum_{i=1}^{m} u_i g_i(x) \left[\frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right]$$

$$p(x,u,\mu) \equiv -M^{-1}\left[b^1 - \mu \sum_{i=1}^{m} \nabla g_i(x)/g_i(x) \right] = -\left(\frac{1}{\mu}M\right)^{-1} \nabla\Phi(x,\mu)$$

When clarity is not affected, let $p \equiv p(x,u,\mu)$, $H \equiv H(x,\mu)$, and $M \equiv M(x,u,\mu)$.

$$\| p \|_H \equiv (p^T H p)^{1/2}$$

$$\| p \|_{M/\mu} \equiv \left(p^T \left(\frac{1}{\mu} M \right) p \right)^{1/2}$$

For some $z \epsilon E^1$ let the region $\mathbb{R}_z \equiv \{x \epsilon E^n \mid f(x) \leq z,\ x \epsilon \mathbb{R}\}$ be bounded with a non-empty interior. Then for some positive integer r we define the analytic center of $\mathbb{R}_z$ as the point in $\mathbb{R}_z$ which solves

$$\max \ (z - f(x))^r \prod_{i=1}^{m} g_i(x).$$

This dissertation has presented two implementations of the Modified Primal-Dual Algorithm (MPDA), demonstrating two ways to show polynomial complexity for the algorithm. In each of these implementations, the analysis assumes a fairly restrictive initial condition: namely, a known initial point (x^0,u^0) and barrier parameter $\mu^0 > 0$ such that the following conditions hold:

$$x^0 \in \mathbb{R}^* \tag{8.1}$$

$$\left\| p(x^0,u^0,\mu^0) \right\|_{M/\mu^0} \leq \tau \tag{8.2}$$

$$\mu^0 \leq 2^{O(L)}. \tag{8.3}$$

Recall that if $u_i^0 g_i(x^0) - \mu = 0$, $i \in I$, then (8.2) is equivalent to $\left\| p(x^0,u^0,\mu^0) \right\|_H \leq \tau$. Also, (8.3) applies only for problem LP, that is, when the objective function is linear and the constraints linear affine. In practice, one may not receive such a conveniently structured problem. Fortunately, it is a straightforward task to meet these conditions given an arbitrary starting point $\hat{x} \in E^n$. Our first concern is finding a feasible value of x.

VIII.1 Feasibility

The following method for achieving feasibility is a version of the Two Phase method often used in linear programming. For a detailed discussion of that

application, see Bazaraa, Jarvis, and Sherali(1990). Given $\hat{x}$, let $\mathcal{V}$ denote the set of indicies of constraints that are non-positive for $\hat{x}$. Now define a set of artificial variables a_j, $j \in \mathcal{V}$, and a variable $z \in E^1$. Let $a_j^0 = -g_j(\hat{x}) + 1$, $j \in \mathcal{V}$, $z^0 = \max(a_j) + 1$, and formulate the following problem FP.

FP: min z

s.t. $g_i(x) \geq 0$, $i \in I - \mathcal{V}$

$g_j(x) + a_j \geq 0$, $j \in \mathcal{V}$

$z \geq a_j$, $j \in \mathcal{V}$

$z \geq -1$

By construction, FP has a solution and $(\hat{x}, a^0, z^0)$ is strictly feasible for FP. Furthermore, FP has a linear objective function and concave quadratic constraints, so the algorithm of den Hertog, Roos, and Terlaky can solve FP in polynomial time. The number of constraints remains O(m), and if we assume $m = O(n)$, the computational complexity is the same as for problem P. In reality, the algorithm may be terminated as soon it obtains a negative value of z. The associated value of x is strictly feasible for P and may be denoted $\tilde{x}$. If the minimum z is non-negative, then P is infeasible. With $\tilde{x} \in \mathbb{R}^*$, we have satisfied (8.1), so we now proceed to satisfying (8.2) and (8.3), known as centering.

Another approach to finding an interior point is found in Fiacco and McCormick(1968, 1990). Let $\hat{x} \in E^n$ be arbitrary, and define the sets $S \equiv \{i \mid g_i(\hat{x}) \leq 0, i \in I\}$ and $T \equiv \{i \mid g_i(\hat{x}) > 0, i \in I\}$. Now minimize

$$U(x,\mu^k) \equiv -\sum_{i \in S} g_i(x) + \mu^k \sum_{i \in T} ln(g_i(x))$$

for $\mu^k \downarrow 0$ as $k \to +\infty$. When in the course of this minimization a new constraint is strictly satisfied, its index is moved from S to T, $U(x,\mu^k)$ is redefined accordingly, and the process is begun again. This continues until all constraints are strictly satisfied or it is shown no interior point exists.

VIII.2 Centering

A straightforward method for initial centering is found in McCormick(1991b), based on the work of Huard(1967) and Fiacco(1979). Given that $\tilde{x} \in \mathbb{R}^*$ (possibly from the application of the feasibility methodology above), define $V^0 \equiv \{x \in E^n \mid f(x) < f(\tilde{x})\} \cap \mathbb{R}^*$. One then finds an approximate solution to the problem

$$\max_{x \in V^0} [-f(x) + f(\tilde{x})] \prod_{i=1}^{m} g_i(x). \tag{8.4}$$

This is equivalent to finding an approximate solution to

$$\min_{x \in V^0} -ln[-f(x) + f(\tilde{x})] - \sum_{i=1}^{m} ln\big(g_i(x)\big). \tag{8.5}$$

Let x^0 approximately minimize (8.5). Then we have

$$\frac{\nabla f(x^0)}{[-f(x^0) + f(\tilde{x})]} - \sum_{i=1}^{m} \frac{\nabla g_i(x^0)}{g_i(x^0)}$$

approximately equal to zero. Now let $\mu^0 \equiv f(\tilde{x}) - f(x^0)$, and we have (x^0, μ^0) that approximately solves

$$\min_{x \in \mathbb{R}^*} \Phi(x,\mu^0).$$

With the (familiar) definition of $u_i^0 \equiv \mu^0/g_i(x^0)$, and given a good enough approximate solution, we have (x^0, u^0, μ^0) which will satisfy (8.2).

Although this two step approach is a standard one and very workable, we now turn to a method which combines the feasibility and centering operations and makes use of a polynomial-time analytical centering algorithm. The method is based on Jarre (1991), which demonstrated the polynomality of a method of analytical centers when applied to our problem P. The presentation here will not include detail on the polynomality of the centering algorithm, which is beyond the scope of this research--readers interested in such detail are better served by reading the original paper. Rather, we show a scheme adapted from Jarre (1991) which allows the application of his algorithm to finding (x^0, u^0, μ^0) satisfying the initial conditions of the MPDA.

VIII.3 Combined Feasibility and Centering

We make the following definitions:

$$d_i \equiv g_i(0) - 1, \; i \in I$$

$$\hat{g}_i(x,\gamma) \equiv g_i(x) - \gamma d_i, \; i \in I$$

$$w_{m+1} \equiv -\sum_{i=1}^{m} \nabla g_i(0)$$

$$\widehat{g}_{m+1}(\mathrm{x},\nu) \equiv w_{m+1}^{\mathrm{T}}\mathrm{x} + \nu.$$

This gives us

$$\widehat{g}_i(0,1) = 1,\ \mathrm{i} \in (1, 2, \ldots, \mathrm{m}+1)$$

$$\sum_{i=1}^{m+1} \frac{\nabla_x \widehat{g}_i(0,1)}{\widehat{g}_i(0,1)} = 0. \tag{8.6}$$

Now let

$$\widehat{\mathbb{R}}(\gamma,\nu) \equiv \{\mathrm{x} \in \mathrm{E}^n \mid \widehat{g}_i(\mathrm{x},\gamma) \geq 0,\ \mathrm{i} \in \mathrm{I},\ \widehat{g}_{m+1}(\mathrm{x},\nu) \geq 0\}.$$

for some fixed pair (γ,ν) such that $\gamma \in \mathrm{E}^1$, $\nu \leq \infty$. Since $\mathbb{R}$ is a bounded convex polytope and $\widehat{\mathbb{R}}(\gamma,\nu)$ has as constraints perturbations of all the constraints forming $\mathbb{R}$, $\widehat{\mathbb{R}}(\gamma,\nu)$ must be bounded (Fiacco and McCormick(1968, 1990), Theorem 24). The point $\mathrm{x} = 0$ is an element of $\widehat{\mathbb{R}}(1,1)$ so $\widehat{\mathbb{R}}(1,1)$ is non-empty. In fact, (8.6) shows that x=0 is the analytical center of $\widehat{\mathbb{R}}(1,1)$. We now make two successive applications of the analytical centering algorithm, the first with a strictly monotonically increasing sequence $\{\nu^k\}\uparrow\infty$ with $\nu^0 = 1$, and the second with a strictly monotonically decreasing sequence $\{\gamma^k\}\downarrow 0$ with $\gamma^0 = 1$. The first application of the algorithm "backs off" the $\mathrm{m}+1^{th}$ constraint until it no longer has any significant effect on the analytic center, and the second application decreases the perturbations to the first m constraints until $\widehat{\mathbb{R}}$ closely approximates $\mathbb{R}$, and

thus the resulting close approximation of the analytical center of $\widehat{\mathbb{R}}$ is a close aproximation of the analytical center of $\mathbb{R}$. In this way, two applications of a polynomial analytical centering algorithm yield a a close aproximation of the analytical center of $\mathbb{R}$, denoted x^0, which implies

$$\sum_{i=1}^{m} \frac{\nabla g_i(x^0)}{g_i(x^0)}$$

is nearly zero. All that remains is to then select a large value of μ^0 such that $\mu^0 \leq 2^{O(L)}$ so that for problem P,

$$\nabla \Phi(x^0, \mu^0) = \frac{b}{\mu^0} - \sum_{i=1}^{m} \frac{\nabla g_i(x^0)}{g_i(x^0)}$$

remains close to zero, and the usual definition of $u_i^0 \equiv \mu^0 / g_i(x^0)$ yields (x^0, u^0, μ^0) satisfying the initial conditions in (8.1) through (8.3).

CHAPTER IX: SUMMARY AND RECOMMENDATIONS FOR FUTURE RESEARCH

IX.1 Summary of Results

We begin this summary by restating the problem to be solved along with important notational conventions.

Standard Notation and Definitions

General Constrained Optimization Problem:

$$\min_{i \in E^n} f(x)$$

$$\text{s.t. } g_i(x) \geq 0, \; i \in I \equiv (1, 2, \ldots, m)$$

$$\mathbb{R} \equiv \{x \in E^n \mid g_i(x) \geq 0, \; i \in I\}$$

$$\mathbb{R}^* \equiv \{x \in E^n \mid g_i(x) > 0,\ i \in I\}$$

$$\Phi(x,\mu) \equiv \frac{f(x)}{\mu} - \sum_{i=1}^{m} \ln(g_i(x)) \quad \text{for } x \in \mathbb{R}^* \text{ and } \mu > 0.$$

$$\nabla\Phi(x,\mu) = \frac{\nabla f(x)}{\mu} - \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)}$$

$$H \equiv \nabla^2\Phi(x,\mu) = \frac{\nabla^2 f(x)}{\mu} + \sum_{i=1}^{m} \frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}$$

(Note: when $f(x)$ is linear, $\nabla^2 f(x)$ vanishes.)

$$L(x,u) \equiv f(x) - \sum_{i=1}^{m} u_i g_i(x)$$

$$\nabla L(x,u) = \nabla f(x) - \sum_{i=1}^{m} u_i \nabla g_i(x) = 0,\ u_i \geq 0,\ i \in I$$

$$M = \sum_{i=1}^{m} u_i g_i(x) \left[\frac{\nabla g_i(x)\nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right] \text{for } x \in \mathbb{R}^* \text{ and } u_i > 0,\ i \in I.$$

$$p \equiv -M^{-1}\left[b^T - \mu \sum_{i=1}^{m} \nabla g_i(x)/g_i(x) \right] = -\left(\frac{1}{\mu}M\right)^{-1} \nabla\Phi(x,\mu)$$

$$\|p\|_H \equiv (p^T H p)^{1/2}$$

$$\|p\|_{M/\mu} \equiv \left(p^T\left(\frac{1}{\mu}M\right)p\right)^{1/2}$$

$$\omega \equiv \max_{i \in I} \left|1 - \frac{u_i g_i(x)}{\mu}\right|$$

$$\sqrt{1-\omega}\,\|p\|_H \leq \|p\|_{M/\mu} \leq \sqrt{1+\omega}\,\|p\|_H \text{ for } \omega < 1 \text{ and } f(x) \text{ linear}$$

This dissertation began with a review of concepts fundamental to constrained optimization with interior point methods and current research in the field. That was followed by a detailed description of the Primal-Dual Algorithm developed by McCormick(1991b). Recall the basic idea of that algorithm is to use Newton's method with a step size line search to find solutions to the following set of equations:

$$\nabla_x L(x,u) = 0$$

$$u_i g_i(x) - \mu^k = 0,\ i \in I$$

where $x \in \mathbb{R}^*$ and $u_i > 0$, $i \in I$. These equations are solved for a sequence of

positive μ^k such that $\{\mu^k\}\downarrow 0$ as $k\rightarrow+\infty$. The first major research result, in Chapter IV, was showing the similarity of p, the primal variable Newton direction from the Primal-Dual Algorithm, with the Newton direction $-H^{-1}\nabla\Phi(x\mu)$ for the standard SUMT with the logarithmic barrier function. This similarity was exploited by constructing a Primal Variable Algorithm using the search direction p, and limiting the divergence of p from $-H^{-1}\nabla\Phi(x\mu)$ through the measure ω. It was shown that by keeping ω small ($\omega \leq 1/62$), the SUMT polynomiality results of den Hertog, Roos, and Terlaky(1990) could be used as a template to construct a convergence and polynomial complexity proof for the Primal Variable Algorithm.

In the Primal Variable Algorithm analysis, $\omega \leq 1/62$ was specified. Ensuring that ω meets that criterion is another matter. The degree to which $u_i g_i(x) - \mu^k = 0$, $i \in I$ is violated during the course of the Primal-Dual Algorithm is very dependent on the way the u_i variables change. Thus, Chapter V contained research into the performance of the dual variable Newton direction for an arbitrary u_i, particularly when using a step size generated by the Primal Variable Algorithm. The result was two cases, determined by the sign of $u_i g_i(x) - \mu^k$. Both cases revealed that the maintenance of $\omega \leq 1/62$ is not certain when movement in the dual variables u_i, $i \in I$, according to the dual Newton direction, is appended to the Primal-Variable Algorithm. Thus, further study was required to establish polynomiality for the Primal-Dual Algorithm.

In Chapter VI, research centered on developing a modification to the dual variable Newton step for an arbitrary u_i to guarantee $\omega \leq 1/62$. First, the simpler case of a linear affine $g_i(x)$ lead to the discovery that a simple step size change

using existing information could guarantee a monotone decrease in $\left|u_i g_i(x) - \mu^k\right|$ with each Newton step. Extending this result to the more general quadratic $g_i(x)$ required both a step size change and an additive term to the dual Newton step, again with existing information. With a monotone decrease in $\left|u_i g_i(x) - \mu^k\right|$ at each Newton step now ensured for $i \in I$, movement in the dual variables according to the modified to the dual Newton step was added to the Primal Variable Algorithm, resulting in the Modified Primal-Dual Algorithm (MPDA). The MPDA is essentially the original Primal-Dual Algorithm with the modified dual Newton step, using $\Phi(x,\mu)$ as a merit function. The modified dual Newton step and small decrements to μ guarantee $\omega \leq 1/62$, so convergence and polynomiality follow from the Primal Variable Algorithm.

The effect of the modification to the dual Newton step had further consequences. In Chapter VII, analysis revealed that when the MPDA is applied to a linear program (LP), the polynomiality results of Anstreicher(1990) apply directly. Anstreicher's work is remarkable for showing polynomiality of the original SUMT of Fiacco and McCormick for quadratic programming (QP) when initial conditions and parameter values are correctly specified. When a linear objective function is specified to conform to our MPDA analysis, QP becomes LP, and it is shown that Anstreicher's proofs extend to the MPDA. Hence, we have another convergence and polynomial complexity result, although for a less general class of problem.

The research of this dissertation concludes in Chapter VIII, which addressed the problem of finding an initial point (x^0,u^0,μ^0) which meets the fairly restrictive

initial conditions used by both den Hertog, Roos, and Terlaky(1990) and Anstreicher(1990). The conditions are feasibility, namely $x^0 \in \mathbb{R}^*$, $u^0 > 0$, and $\mu^0 \geq 0$; and centering, meaning $\left\| p(x^0,u^0,\mu^0) \right\|_H$ is small. Techniques to achieve feasibility are covered by many authors--feasibility results from Bazaara, Jarvis, and Sherali(1990) and Fiacco(1979) apply, and were covered in Chapter VIII. The chapter also presented a centering method from McCormick(1991b), based on the work of Huard(1967) and Fiacco(1979), and concluded with a combined feasibility and centering technique adapted from Jarre(1991).

IX.2 Future Research

It is possible that the results developed here for convex programming with quadratic constraints could be extended to smooth convex programming. We define smooth convex programming by way of a relative Lipschitz condition set forth in Jarre(1991). Let x and $x + h$ in $\mathbb{R}^*$ and $z \in E^n$ be arbitrary. Then we say the relative Lipschitz condition is satisfied if for some $M \geq 0$, the Hessian of each $g_i(x)$ satisfies the following:

$$\left| z^T \left[\nabla^2 g_i(x+h) - \nabla^2 g_i(x) \right] z \right| \leq M \| h \|_{H(x)} z^T \nabla^2 g_i(x) z$$

where

$$H(x) \equiv \sum_{i=1}^{m} \frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)}.$$

The convergence and complexity results for the logarithmic barrier function in den Hertog, Roos, and Terlaky(1990) extend to smooth convex programs, with M entering into the complexity polynomial. It should be straightforward to make the same extension to the Modified Primal-Dual Algorithm.

Another research area is the choice of merit function. For the general constrained optimization problem

$$\min_{i \in E^n} f(x)$$

$$\text{s.t. } g_i(x) \geq 0,\ i \in I,$$

there are two candidate functions which incorporate primal and dual variables mentioned in McCormick(1991a),

$$MF1(x,u,\mu) \equiv f(x) - \mu \sum_{i=1}^{m} ln(g_i(x)) + \sum_{i=1}^{m} \left[u_i g_i(x) - \mu ln(u_i g_i(x))\right], \text{ and}$$

$$MF2(x,u,\mu) \equiv \nabla L(x,u)^T M^{-1} \nabla L(x,u) + \sum_{i=1}^{m} \frac{1}{\mu} [u_i g_i - \mu]^2,$$

for $x \in \mathbb{R}^*$, $u > 0$, and $\mu > 0$, where the matrix M is defined as in earlier chapters:

$$M = \sum_{i=1}^{m} u_i g_i(x) \left[\frac{\nabla g_i(x) \nabla g_i(x)^T}{g_i(x)^2} - \frac{\nabla^2 g_i(x)}{g_i(x)} \right] \text{ for } x \in \mathbb{R}^* \text{ and } u_i > 0,\ i \in I.$$

MF1(x,u,μ) has the interesting property in that a stationary point with respect to x and u for MF1(x,u,μ) satisfies the equations motivating the Primal-Dual Algorithm (McCormick(1991b)):

$$\nabla L(x,u) = 0$$

$$u_i g_i(x) - \mu = 0,\ i \in I.$$

Successful use of such a merit function might reduce or eliminate the need for a modification to the Primal-Dual Algorithm to show polynomiality.

Finally, only limited mention has been made in this dissertation to convergence of the Lagrangian multipliers. Further research into their behavior, particularly in the presence of degeneracy or alternate optimal solutions, is in order. Within the context of quadratic programming, alternate optimal solutions are possible whenever the objective function has a singular Hessian, that is, the Hessian is not positive definite.

APPENDIX 1

This appendix contains the proofs for the lemmas and theorems from Anstreicher, which are presented without proof in Chapter 7.

Lemma VII.1: Let x be feasible in QP and $(\hat{x}, \hat{u})$ be feasible in QD. Then $f(x) - d(\hat{x}, \hat{u}) \geq 0$.

Proof: $f(x) - d(\hat{x}, \hat{u}) = b^T x + \frac{1}{2}x^T Q x + \frac{1}{2}\hat{x}^T Q \hat{x} - c^T \hat{u}$. Since $(\hat{x}, \hat{u})$ is dual feasible, $b^T = \hat{u}^T A - \hat{x}^T Q$, $\Rightarrow$

$$f(x) - d(\hat{x}, \hat{u}) = \tfrac{1}{2}x^T Q x + \hat{u}^T A x - \hat{x}^T Q x + \tfrac{1}{2}\hat{x}^T Q \hat{x} - \hat{u}^T c$$

$$= \hat{u}^T(Ax - c) + \tfrac{1}{2}x^T Q x - \hat{x}^T Q x + \tfrac{1}{2}\hat{x}^T Q \hat{x}$$

Now define $s \equiv Ax - c$.

$$f(x) - d(\hat{x}, \hat{u}) = \hat{u}^T s + \tfrac{1}{2}x^T Q x - \hat{x}^T Q x + \tfrac{1}{2}\hat{x}^T Q \hat{x}$$

$$= \hat{u}^T s + \tfrac{1}{2}(x^T - \hat{x}^T)Qx + \tfrac{1}{2}\hat{x}^T Q(\hat{x} - x)$$

$$= \hat{u}^T s + \tfrac{1}{2}(x - \hat{x})^T Qx - \tfrac{1}{2}(x - \hat{x})^T Q\hat{x}$$

$$= \hat{u}^T s + \tfrac{1}{2}(x - \hat{x})^T Q(x - \hat{x})$$

Thhe feasibility of x in QP implies $s \geq 0$. We also have $\hat{u} \geq 0$ and Q positive definite. Thus

$$f(x) - d(\hat{x}, \hat{u}) = \hat{u}^T s + \tfrac{1}{2}(x - \hat{x})^T Q(x - \hat{x}) \geq 0$$

which proves the lemma.

Lemma VII.2: Let $x \in \mathbb{R}^*$ and $\mu > 0$ be arbitrary. Then
$\left\| S^{-1} p_s \right\|^2 \leq \| p_x \|_H^2 \leq \delta(x,\mu)^2$.

Proof: $\| p_x \|_H^2 = p_x^T \left(\frac{1}{\mu} Q + A^T S^{-2} A \right) p_x \geq p_x^T (A^T S^{-2} A) p_x = \left\| S^{-1} p_s \right\|^2$, for the first inequality.

Now let $u = u(x,\mu)$ which we define as the minimizing value of u resulting in $\delta(x,\mu)$. Therefore $A^T u = b + Qx$, and

$$\| p_x \|_H^2 = \nabla\Phi(x,\mu)^T H^{-1} \nabla\Phi(x,\mu)$$

$$= \left(\frac{Qx+b}{\mu} - A^T s^{-1}\right)^T \left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} \left(\frac{Qx+b}{\mu} - A^T s^{-1}\right)$$

$$= \left(A^T\left(\frac{u}{\mu} - s^{-1}\right)\right)^T \left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} \left(A^T\left(\frac{u}{\mu} - s^{-1}\right)\right)$$

$$= \left(\frac{Su}{\mu} - e\right)^T S^{-1} A\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1} \left(\frac{Su}{\mu} - e\right)$$

$$\leq \sigma \left\| \frac{Su}{\mu} - e \right\|^2 = \sigma \delta(x,\mu)^2$$

where σ is the maximum eigenvalue of

$$S^{-1} A\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1}. \qquad \text{(A1.1)}$$

To evaluate σ, we see that

$$\left[S^{-1} A\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1}\right]^2$$

$$= S^{-1} A\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1} S^{-1} A\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1}$$

$$= S^{-1} A\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} (A^T S^{-2} A)\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1} \qquad \text{(A1.2)}$$

The maximum eigenvalue of (A1.2) is less than or equal to that of the following matrix (A1.3)

$$S^{-1} A\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} \left(\frac{1}{\mu}Q + A^T S^{-2} A\right)\left(\frac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1} \qquad \text{(A1.3)}$$

$$= S^{-1}A\left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1}$$

which is (A1.1). Since the maximum eigenvalue of the square of (A1.1) is less than that of (A1.1), namely σ, we must have $\sigma \leq 1$, which implies $\| p_x \|_H^2 \leq \delta(x,\mu)^2$ and completes the proof of the lemma.

Lemma VII.3: Given $x \in \mathbb{R}^*$, suppose $\left\| S^{-1} p_s \right\| < 1$. Let $\bar{x} = x + p_x$, $\bar{s} = s + p_s$, and $\bar{S} = \text{diag}(\bar{s})$. Let $\bar{H} = \nabla^2 \Phi(\bar{x}, \mu)$, $p_{\bar{x}} = -\bar{H}^{-1} \nabla \Phi(\bar{x}, \mu)$, and $p_{\bar{s}} = A p_{\bar{x}}$. Then $\bar{x} \in \mathbb{R}^*$ and

$$\left\| \bar{S}^{-1} p_{\bar{s}} \right\| \leq \| p_{\bar{x}} \|_{\bar{H}}^2 \leq \delta(\bar{x}, \mu) \leq \left\| S^{-1} p_s \right\|^2 \leq \| p_x \|_H^2 \leq \delta(x,\mu)^2$$

Proof: $\bar{s} = s + p_s = S\left(e + S^{-1} p_s\right)$. Since $\left\| S^{-1} p_s \right\| < 1$, the absolute value of each element of the vector $S^{-1} p_s$ must be less than 1, and each element of $\left(e + S^{-1} p_s\right)$ is strictly positive. The vector s is also strictly positive, so $\bar{s}$ must be as well and therefore $\bar{x} \in \mathbb{R}^*$. Now consider

$$\delta(\bar{x}, \mu) = \min_{u \mid A^T u = b + Q\bar{x}} \left\| \tfrac{1}{\mu} \bar{S} u - e \right\|$$

Let u satisfy $A^T u = b + Qx$. (One such value is $u = u(x,\mu)$.) Then

$$\bar{x} = x - H^{-1} \nabla \Phi(x,\mu)$$

$$= x - \left(\tfrac{1}{\mu} Q + A^T S^{-2} A\right)^{-1} \left(\frac{Qx + b}{\mu} - A^T s^{-1}\right)$$

$$= x - \left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1}\left(\frac{Su}{\mu} - e\right)$$

Some manipulation yields

$$\tfrac{1}{\mu}Q(x - \bar{x}) = \tfrac{1}{\mu}Q\left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1}\left(\frac{Su}{\mu} - e\right)$$

and using the fact that $\tfrac{1}{\mu}Q\left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} = I - A^T S^{-2} A\left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1}$ we obtain

$$\tfrac{1}{\mu}Q(x - \bar{x}) = A^T S^{-1}\left(\frac{Su}{\mu} - e\right) - A^T S^{-2} A\left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1} A^T S^{-1}\left(\frac{Su}{\mu} - e\right).$$

$$= A^T S^{-1}\left(\frac{Su}{\mu} - e\right) - A^T S^{-2} A\left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1}\left(\frac{A^T u}{\mu} - A^T s^{-1}\right)$$

$$= A^T S^{-1}\left(\frac{Su}{\mu} - e\right) - A^T S^{-1}(S^{-1} A)\left(\tfrac{1}{\mu}Q + A^T S^{-2} A\right)^{-1}\left(\frac{b + Qx}{\mu} - A^T s^{-1}\right)$$

$$= A^T S^{-1}\left(\frac{Su}{\mu} - e\right) - A^T S^{-1}(S^{-1} A) H^{-1} \nabla \Phi(x, \mu)$$

$$= A^T S^{-1}\left(\frac{Su}{\mu} - e\right) + A^T S^{-1}\left(S^{-1} A p_x\right)$$

$$= A^T S^{-1}\left(\frac{Su}{\mu} - e + S^{-1} p_s\right) \qquad \text{(A1.4)}$$

From (A1.4) we see that

$$Q\bar{x} = Qx - A^T u + \mu A^T S^{-1}\left(e - S^{-1} p_s\right)$$

$$= -b + \mu A^T S^{-1}\left(e - S^{-1}p_s\right) \qquad \text{(A1.5)}$$

so by setting $\bar{u} = \mu S^{-1}\left(e - S^{-1}p_s\right)$ we have $A^T\bar{u} = b + Q\bar{x}$. Now let $P_s = \text{diag}(p_s)$ and let $\left(S^{-1}p_s\right)^2$ denote the vector of the squares of the elements of $S^{-1}p_s$. Then

$$\delta(\bar{x},\mu) = \min_{u|\ A^T u = b + Q\bar{x}} \left\| \tfrac{1}{\mu}\bar{S}\, u - e \right\|$$

$$\leq \left\| \tfrac{1}{\mu}\bar{S}\,\bar{u} - e \right\|$$

$$= \left\| (S + P_s)\, S^{-1}\left(e - S^{-1}p_s\right) - e \right\|$$

$$= \left\| \left(I + S^{-1}P_s\right)\left(e - S^{-1}p_s\right) - e \right\|$$

$$= \left\| e - S^{-1}p_s + S^{-1}p_s - S^{-1}P_s S^{-1}p_s - e \right\|$$

$$= \left\| \left(S^{-1}p_s\right)^2 \right\|$$

$$\leq \left\| S^{-1}p_s \right\|^2$$

This proves $\delta(\bar{x},\mu) \leq \left\| S^{-1}p_s \right\|^2$, and the remaining inequalities are a result of Lemma VII.2.

Lemma VII.4: If $x \in \mathbb{R}^*$ and $\|p_x\|_H < 1$, then $\Phi(x,\mu) - \Phi(x(\mu),\mu) \leq \dfrac{\|p_x\|_H^2}{1 - \|p_x\|_H^2}$.

Proof: For an arbitrary fixed $\mu > 0$, $\Phi(\cdot,\mu)$ is convex, so

$$\Phi(x,\mu) - \Phi(x + p_x,\mu) \leq -\nabla\Phi(x,\mu)^T p_x = \|p_x\|_H^2 \tag{A1.6}$$

Let $x^0 = x$, and define $\{x^0, x^1, x^2, \ldots\}$ as the sequence of points obtained by generating a succession of Newton steps starting at x. From (A1.6) and Lemma VII.3,

$$\Phi(x,\mu) - \Phi(x(\mu),\mu) = \sum_{i=0}^{\infty}\Big(\Phi(x^i,\mu) - \Phi(x^{i+1},\mu)\Big)$$

$$\leq \sum_{i=0}^{\infty}\|p_x\|_H^{2^{i+1}}$$

$$\leq \frac{\|p_x\|_H^2}{1 - \|p_x\|_H^2}$$

thus proving the lemma.

Lemma VII.5: If $x \in \mathbb{R}^*$ and $\|p_x\|_H < 1$, then

$$\big|f(x) - f(x(\mu))\big| \leq \frac{\|p_x\|_H(1 + \|p_x\|_H)}{1 - \|p_x\|_H}\mu\sqrt{m}$$

Proof: From the convexity of $f(\cdot)$ we get

$$\nabla f(x)^T p_x \leq f(x + p_x) - f(x) \leq \nabla f(x + p_x)^T p_x. \tag{A1.7}$$

If the right hand and left hand sides of (A1.7) can be appropriately bounded, the lemma can be proved. Beginning with the left hand side and the fact that $\nabla f(x) = Qx + b = \mu\nabla\Phi(x,\mu) + \mu A^T S^{-1} e$ we see

$$\nabla f(x)^T p_x = \mu\nabla\Phi(x,\mu)^T p_x + \mu e^T S^{-1} A p_x$$

$$= \mu\nabla\Phi(x,\mu)^T p_x + \mu e^T S^{-1} p_s. \tag{A1.8}$$

Now $\nabla\Phi(x,\mu)^T p_x = \nabla\Phi(x,\mu)^T\Big(- H^{-1}\nabla\Phi(x,\mu)\Big) = - \nabla\Phi(x,\mu)^T H^{-1} H H^{-1}\nabla\Phi(x,\mu)$ $= - \| p_x \|_H^2$. From the Cauchy-Schwartz Inequality and Lemma VII.2 we have

$$e^T S^{-1} p_s \geq - \left| e^T S^{-1} p_s \right| \geq - ||e|| \left\| S^{-1} p_s \right\| \geq - \sqrt{m}\, \| p_x \|_H .$$

Substitution into (A1.8) yields

$$\nabla f(x)^T p_x \geq - \mu \| p_x \|_H^2 - \mu\sqrt{m}\, \| p_x \|_H . \tag{A1.9}$$

Using (A1.5) and noting that $\bar{x} = x + p_x$, we can bound the right hand side of (A1.7) as follows:

$$\nabla f(x + p_x)^T p_x = (Q\bar{x} + b)^T p_x$$

$$= \mu p_x^T A^T S^{-1}\Big(e - S^{-1} p_s \Big)$$

$$= \mu e^T\left(S^{-1}Ap_x\right) - \mu p_x^T A^T S^{-1} S^{-1} p_s$$

$$= \mu e^T\left(S^{-1}p_s\right) - \mu\left\|S^{-1}p_s\right\|^2$$

$$\leq \mu\sqrt{m}\,\|p_x\|_H. \tag{A1.10}$$

We can now subsitute (A1.9) and (A1.10) into (A1.8) to achieve

$$-\mu\|p_x\|_H^2 - \mu\sqrt{m}\,\|p_x\|_H \leq f(x+p_x) - f(x) \leq \mu\sqrt{m}\,\|p_x\|_H$$

$$-\mu\sqrt{m}\,\|p_x\|_H^2 - \mu\sqrt{m}\,\|p_x\|_H \leq f(x+p_x) - f(x) \leq \mu\sqrt{m}\,\|p_x\|_H + \mu\sqrt{m}\,\|p_x\|_H^2$$

$$\left|f(x+p_x) - f(x)\right| \leq \mu\sqrt{m}\,\|p_x\|_H(1+\|p_x\|_H) \tag{A1.11}$$

Again let $x^0 = x$ and $\{x^0, x^1, x^2, \ldots\}$ denote the sequence of points obtained by generating a succession of Newton steps starting at x. Applying (A1.11) and Lemma VII.3 we have

$$\left|f(x) - f(x(\mu))\right| = \left|\sum_{i=0}^{\infty}\left(f(x^i) - f(x^{i+1})\right)\right|$$

$$\leq \sum_{i=0}^{\infty}\left|f(x^i) - f(x^{i+1})\right|$$

$$\leq \sum_{i=0}^{\infty}\mu\sqrt{m}\,\|p_x\|_H^{2^i}\left(1+\|p_x\|_H^{2^i}\right).$$

Since $(1+\|p_x\|_H) \geq \left(1+\|p_x\|_H^{2^i}\right)$ for all $i \geq 0$ this becomes

$$|f(x)-f(x(\mu))| \leq \mu\sqrt{m}(1+\|p_x\|_H)\sum_{i=0}^{\infty}\|p_x\|_H^{2^i}$$

$$\leq \frac{\|p_x\|_H(1+\|p_x\|_H)}{1-\|p_x\|_H}\mu\sqrt{m}$$

which proves the lemma.

Theorem VII.1: Let z^* denote the minimum value of f(x) for problem QP. After $O(L/\theta)$ outer iterations, the algorithm obtains $\mu = \mu^k$ such that $m\mu^k \leq 2^{-O(L)}$ and a point $x = x^{k+1}$ such that $f(x) - z^* \leq \left(1+\frac{\tau(1+\tau)}{(1-\tau)\sqrt{m}}\right)2^{-O(L)}$.

Proof: The duality gap at $x(\mu^k)$ is $m\mu^k$, and at the end of the k^{th} outer iteration we have $\|p_x\|_H \leq \tau$. This along with Lemma VII.5 gives us

$$f(x)-z^* = f(x(\mu^k))-z^*+f(x)-f(x(\mu^k)) \leq m\mu^k + \frac{\tau(1+\tau)}{(1-\tau)}\sqrt{m}\mu^k$$

$$\leq m\mu^k\left(1+\frac{\tau(1+\tau)}{(1-\tau)\sqrt{m}}\right) \qquad \text{(A1.12)}$$

To show $m\mu^k \leq 2^{-O(L)}$, note that $m\mu^k = m(1-\theta)^k\mu^0$, and $\mu^0 \leq 2^{O(L)}$ is assumed. So we need a value of k such that

$$m(1-\theta)^k 2^{O(L)} \leq 2^{-O(L)}$$

Taking logarithms,

$$k ln(1-\theta) + O(L) + ln(m) \leq -O(L)$$

$$k \geq \frac{O(L) + ln(m)}{-ln(1-\theta)}$$

and since $ln(m) \leq O(L)$ and $\theta \leq -ln(1-\theta)$, we obtain

$$k \geq O\left(\frac{L}{\theta}\right)$$

which completes the proof.

Lemma VII.6: Let $\overline{\lambda} = \frac{1}{1+\|p_x\|_H}$. Then

$$\Phi(x,\mu) - \Phi(x + \overline{\lambda}\, p_x, \mu) \geq \|p_x\|_H - ln(1+\|p_x\|_H)$$

Proof: The Taylor series expansion for $\Phi(x + \lambda p_x, \mu)$ is

$$\Phi(x + \lambda p_x, \mu) = \Phi(x,\mu) + \lambda \nabla\Phi(x,\mu)^T p_x + \tfrac{1}{2}\lambda^2 p_x^T H p_x + \sum_{j=3}^{\infty} t_j \qquad \text{(A1.13)}$$

where t_j denotes the j^{th} term of the Taylor expansion. For $j \geq 3$, the j^{th} term may be computed as

$$t_j = \frac{(-\lambda)^j}{j} \sum_{i=1}^{m} \left(S^{-1}p_s\right)_i^j$$

where $\left(S^{-1}p_s\right)_i^j$ denotes the j^{th} power of the i^{th} component of $S^{-1}p_s$. Thus

$$|t_j| \leq \frac{\lambda^j}{j} \sum_{i=1}^{m} \left|\left(S^{-1}p_s\right)_i^j\right| \leq \frac{\lambda^j}{j} \left(\sum_{i=1}^{m} \left|\left(S^{-1}p_s\right)_i\right| \right)^{j/2} = \frac{\lambda^j}{j} \left\|S^{-1}p_s\right\|^j$$

and by Lemma VII.2,

$$|t_j| \leq \frac{\lambda^j}{j} \|p_x\|_H^j. \tag{A1.14}$$

Expanding the linear and quadratic terms of the Taylor series and again using Lemma VII.2, we obtain

$$\lambda \nabla\Phi(x,\mu)^T p_x + \tfrac{1}{2}\lambda^2 p_x^T H p_x = -\lambda \nabla\Phi(x,\mu)^T H^{-1} H H^{-1} \nabla\Phi(x,\mu) + \tfrac{1}{2}\lambda^2 p_x^T H p_x$$

$$= \left(-\lambda + \tfrac{1}{2}\lambda^2\right) \|p_x\|_H^2. \tag{A1.15}$$

Substituting (A1.14) and (A1.15) into (A1.13) yields

$$\Phi(x + \lambda p_x, \mu) \leq \Phi(x,\mu) + \left(-\lambda + \tfrac{1}{2}\lambda^2\right) \|p_x\|_H^2 + \sum_{j=3}^{\infty} \frac{\lambda^j}{j} \|p_x\|_H^j. \tag{A1.16}$$

Furthermore, the Taylor series expansion of $-ln\left(1 - \lambda\|p_x\|_H\right)$ is

$$-ln\left(1 - \lambda\|p_x\|_H\right) = -ln(1) + \lambda\|p_x\|_H + \tfrac{1}{2}\left(\lambda\|p_x\|_H\right)^2 + \tfrac{1}{3}\left(\lambda\|p_x\|_H\right)^3 + \tfrac{1}{4}\left(\lambda\|p_x\|_H\right)^4 + \ldots$$

which implies

$$\sum_{j=3}^{\infty} \frac{\lambda^j}{j} \| p_x \|_H^j = -ln\Big(1 - \lambda \| p_x \|_H\Big) - \lambda \| p_x \|_H - \frac{1}{2}\Big(\lambda \| p_x \|_H\Big)^2 \qquad \text{(A1.17)}$$

Now we substitute (A1.17) into (A1.16) and find

$$\Phi(x + \lambda p_x, \mu) \leq \Phi(x, \mu) - \lambda \| p_x \|_H^2 - ln\Big(1 - \lambda \| p_x \|_H\Big) - \lambda \| p_x \|_H.$$

The right hand side is minimized by $\lambda = \overline{\lambda} = \frac{1}{1 + \| p_x \|_H}$, so

$$\Phi(x, \mu) - \Phi(x + \overline{\lambda}\, p_x, \mu) \geq \overline{\lambda} \| p_x \|_H (1 + \| p_x \|_H) - ln\Big(1 - \overline{\lambda} \| p_x \|_H\Big)$$

$$= \| p_x \|_H - ln(1 + \| p_x \|_H)$$

which proves the lemma.

Corollary VII.6: Let $\tau < \| p_x \|_H \leq \frac{1}{2}$. Then a pure Newton step ($\lambda = 1$) obtains

$$\Phi(x, \mu) - \Phi(x + p_x, \mu) \geq \Delta = \tau^2 + ln(1 - \tau) + \tau > 0$$

Proof: Using a step size of $\lambda = 1$, from Lemma VII.6 we have

$$\Phi(x + p_x, \mu) \leq \Phi(x, \mu) - \| p_x \|_H^2 - ln(1 - \| p_x \|_H) - \| p_x \|_H$$

or equivalently

$$\Phi(x,\mu) - \Phi(x + p_x,\mu) \geq \|p_x\|_H^2 + ln(1 - \|p_x\|_H) + \|p_x\|_H. \qquad \text{(A1.18)}$$

The right hand side of (A1.18) is positive and increasing for $0 < \|p_x\|_H \leq \frac{1}{2}$, so $0 < \tau < \|p_x\|_H \leq \frac{1}{2}$ implies

$$0 < \Delta = \tau^2 + ln(1 - \tau) + \tau \leq \Phi(x,\mu) - \Phi(x + p_x,\mu)$$

which completes the corollary.

Theorem VII.2: Let $\Delta = \tau - ln(1 + \tau)$, $E_\Phi = \frac{\tau^2}{1 - \tau^2}$, $E_f = \frac{\tau(1 + \tau)}{1 - \tau}$, and N denote the number of inner iterations required for an arbitrary outer iteratation. Then N is bounded as:

$$N \leq \frac{1}{\Delta}\left(\frac{\theta}{1 - \theta}\left(\theta m + 2E_f\sqrt{m}\right) + E_\Phi\right)$$

Proof: Anstreicher's proof is a generalization of a proof by Gonzaga for the linear case. The $(k + 1)^{st}$ outer iteration begins with the point (x^k,μ^k) and $\left\|p(x^k,\mu^{k-1})_x\right\|_H \leq \tau$ for $H = H(x^k,\mu^{k-1})$. Let N denote the number of inner iterations. Lemma VII.6 shows that each inner iteration will decrease $\Phi(\cdot,\mu^k)$ by at least Δ. Following the N iterations, we have x^{k+1} with $\left\|p(x^{k+1},\mu^k)_x\right\|_H \leq \tau$ for $H = H(x^{k+1},\mu^k)$. So

$$\Delta N \leq \Phi(x^k,\mu^k) - \Phi(x^{k+1},\mu^k). \qquad \text{(A1.19)}$$

The right hand side can be bounded. From the definition of $\Phi(x,\mu)$,

$$\Phi(x,\mu^k) = \Phi(x,\mu^{k-1}) + \frac{f(x)}{\mu^k} - \frac{f(x)}{\mu^{k-1}}$$

$$= \Phi(x,\mu^{k-1}) + \left(\frac{\theta}{1-\theta}\right)\frac{f(x)}{\mu^{k-1}}$$

and this leads to

$$\Phi(x^k,\mu^k) - \Phi(x^{k+1},\mu^k) = \Phi(x^k,\mu^{k-1}) - \Phi(x^{k+1},\mu^{k-1})$$

$$+\left(\frac{\theta}{1-\theta}\right)\frac{1}{\mu^{k-1}}\left(f(x^k) - f(x^{k+1})\right). \qquad \text{(A1.20)}$$

Analyzing $\left(f(x^k) - f(x^{k+1})\right)$, we use the the fact that x^k and x^{k+1} are approximately centered with respect to μ^{k-1} and μ^k respectively, and Lemma VII.5 to obtain

$$f(x^k) - f(x^{k+1}) \le f(x(\mu^{k-1})) + E_f\mu^{k-1}\sqrt{m} \; - f(x(\mu^k)) + E_f\mu^k\sqrt{m}$$

$$= f(x(\mu^{k-1})) - f(x(\mu^k)) + E_f(2-\theta)\mu^{k-1}\sqrt{m}.$$

Recall the dual objective function $d(x,u)$ from Lemma VII.1, and $u(x,\mu)$ from Lemma VII.2. Since $u(x(\mu),\mu)$ is dual feasible, with the monotonicity with respect to μ of the dual objective function we have

$$f(x^k) - f(x^{k+1}) \le f(x(\mu^{k-1})) - d(x(\mu^{k-1})),\ u(x(\mu^{k-1}),\mu^{k-1})$$

$$+\ d(x(\mu^k),\ u(x(\mu^k),\mu^k)) - f(x(\mu^k)) + E_f(2-\theta)\mu^{k-1}\sqrt{m}$$

$$\leq f(x(\mu^{k-1})) - d(x(\mu^{k-1}),\ u(x(\mu^{k-1}),\mu^{k-1}))$$

$$-\Big(f(x(\mu^k)) - d(x(\mu^k),\ u(x(\mu^k),\mu^k))\Big) + 2E_f\mu^{k-1}\sqrt{m}$$

$$\leq m\mu^{k-1} - m\mu^k + 2E_f\mu^{k-1}\sqrt{m}$$

$$\leq m\mu^{k-1} - m\mu^{k-1} + \theta m\mu^{k-1} + 2E_f\mu^{k-1}\sqrt{m}$$

$$\leq \mu^{k-1}\Big(\theta m + 2E_f\sqrt{m}\Big). \qquad \text{(A1.21)}$$

We can now use Lemma VII.4 with $\|p_x\|_H \leq \tau$, x^k nearly centered with respect to μ^{k-1}, and $x(\mu^{k-1})$ minimizing $\Phi(x,\mu^{k-1})$ to achieve

$$\Phi(x^k,\mu^{k-1}) - \Phi(x^{k+1},\mu^{k-1}) = \Phi(x^k,\mu^{k-1}) - \Phi(x(\mu^{k-1}),\mu^{k-1})$$

$$+ \Phi(x(\mu^{k-1}),\mu^{k-1}) - \Phi(x^{k+1},\mu^{k-1})$$

$$\leq \Phi(x^k,\mu^{k-1}) - \Phi(x(\mu^{k-1}),\mu^{k-1})$$

$$\leq E_\Phi \qquad \text{(A1.22)}$$

Finally, substitution of (A1.21) and (A1.22) into (A1.20) yields

$$\Phi(x^k,\mu^k) - \Phi(x^{k+1},\mu^k) \leq E_\Phi + \left(\frac{\theta}{1-\theta}\right)\frac{1}{\mu^{k-1}}\mu^{k-1}\left(\theta m + 2E_f\sqrt{m}\right)$$

$$\leq \frac{\theta}{1-\theta}\left(\theta m + 2E_f\sqrt{m}\right) + E_\Phi$$

and substitution of this inequality into (A1.19) proves the theorem.

BIBLIOGRAPHY

[1] Anstreicher, K.. 1990. "On long step path following and SUMT for linear and quadratic programming," Technical Paper, Department of Operations Research, Yale University, New Haven, Connecticut 06520.

[2] Anstreicher, K., D. den Hertog, C. Roos, and T. Terlaky. 1990. "A long step barrier method for convex quadratic programming," Report 90-53, Faculty of Mathematics and Informatics/Computer Science, Delft University of Technology, Delft, Netherlands.

[3] Bazaraa, M. S., J. J. Jarvis, and H. D. Sherali. 1990. Linear Programming and Network Flows, John Wiley and Sons, New York.

[4] Charnes, A., and K. O. Kortanek. 1965. "Extreme Point Solutions in Mathematical Programming: An Opposite Sign Algorithm," Systems Research Memorandum No. 129, Northwestern University, Evanston, Illinois.

[5] den Hertog, D., C. Roos, and T. Terlaky. 1990. "On the classical logarithmic barrier function method for a class of smooth convex programming problems," Report 90-28, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands.

[6] Fiacco, A. V., 1979. "Barrier Methods for Nonlinear Programming," in Operations Research Support Methodology, (Industrial Engineering Series Volume 2), A. G. Holzman, (Ed.), Marcel Dekker, Inc., New York, pp. 377-440.

[7] Fiacco, A. V., and G. P. McCormick. 1968. Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, New York. An unabridged corrected re-publication was published by SIAM in the series Classics in Applied Mathematics (1990).

[8] Gonzaga, C. C., 1992. "Path following methods for linear programming," SIAM Review, Vol. 34, No. 2, pp. 167-224.

[9] Huang, S., and K. O. Kortanek. 1991. "On a potential reduction algorithm for LP with simultaneous primal-dual updating," Operations Research Letters 10, pp. 501-507.

[10] Huard, P., 1967. "Resolution of mathematical programming with nonlinear constraints by the method of centers," in Nonlinear Programming, J. Abadie (Ed.), North Holland, Amsterdam.

[11] Jarre, F., 1991. "On the convergence of the method of analytical centers when applied to convex quadratic programs," Mathematical Programming 49, pp. 341-358.

[12] Karmakar, N., 1984. "A New Polynomial-time Algorithm for Linear Programming," Combinatoria, 4, pp. 373-395.

[13] Khachian, L. G., 1979. "A Polynomial Algorithm In Linear Programming," Soviet Mathematics Doklady, 20, pp. 191-194.

[14] Klee, V., and G. J. Minty, 1972. "How Good is the Simplex Algorithm?" in Inequalities III, O. Shisha, (Ed.), Academic, New York, pp. 159-175.

[15] McCormick, G. P., 1983. Nonlinear Programming: Theory, Algorithms, and Applications, John Wiley and Sons, New York.

[16] McCormick, G. P., 1991a. "Resolving the shell dual with a nonlinear primal-dual algorithm," T-545/91, Department of Operations Research, The George Washington University, Washington, DC 20052.

[17] McCormick, G. P., 1991b. "The superlinear convergence of a nonlinear primal-dual algorithm," T-550/91, Department of Operations Research, The George Washington University, Washington, DC 20052.

[18] Mehrotra, S.. 1990. "On finding a vertex solution using interior point methods," Technical Report 89-22, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208.

[19] Mehrotra, S., and J. Sun. 1990. "An algorithm for convex quadratic programming that requires $O(n^{3.5}L)$ arithmetic operations," Mathematics of Operations Research, Vol. 15, No. 2, pp. 342-363.

[20] Papadimitriou, C. H., and K. Steiglitz. 1982. Combinatorial Optimization: Algorithms and Complexity, Prentice-Hall, Inc., Engelwood Cliffs, New Jersey.

[21] Powell, M. J. D.. 1991. "On the number of iterations of Karmarkar's algorithm for linear programming," Numerical Analysis Report 23, Department of Applied Mathmatics and Theoretical Physics, University of Cambridge, Cambridge, England.

[22] Wright, M. H.. 1992. "Interior methods for constrained optimization," in Acta Numerica, Vol. 1, A. Iserles (Ed.), Cambridge University Press, New York, pp. 341-407.